SYNTHESIS AND CONFORMATIONAL STUDIES OF A PROTEIN

by

Thomas W. Muir

A Thesis submitted for the
degree of Doctor of Philosophy

University of Edinburgh
October 1992

This thesis is submitted in part fulfillment of the requirements of the degree of Doctor of Philosophy in the University of Edinburgh. Unless otherwise stated the work described is original and has not been previously submitted, in whole or in part, for any degree at this or any other university.

"Some truths there are so near and obvious that a man need only open up his eyes to see them. Such I take this important one to be, viz. that all the choir of heaven and furniture of the earth, in a word all those bodies which compose the mighty frame of the world, have not any subsistence without a mind - that their being is to be perceived or known" - Berkley.

To my parents

ACKNOWLEDGEMENTS

I would like to thank Professor R. Ramage for the provision of research facilities and to express my gratitude for his advice and constant support throughout the course of this work.

I wish to thank Mr. K. Shaw for his assistance in the technical aspects of the solid phase synthetic work described herein and Mr. B. Whigham for performing amino acid analysis. I am also indebted to those involved in the departmental analytical and spectroscopic services for their prompt and competant work.

I would like to express my appreciation of the contributions made by Dr's J.A. Parkinson and D.W. Thomas in both the acquisition and analysis of NMR spectra.

Special thanks are due to Dr's L. Sawyer and S.M. Fawcett for their advice on protein crystallography and for performing X-ray diffraction on synthetic ubiquitin crystals.

Grateful acknowledgement is made to Professor J. Mayer, Dr. J. Arnold and Mr. L. Layfield of the Biochemistry Dept., University of Nottingham for performing biological assays on synthetic protein samples.

I wish to thank Dr's D.W. Thomas and S. Love for proof-reading this thesis.

I would like to thank the Faculty of Science, University of Edinburgh for the provision of a research grant.

Finally, I would like to express my gratitude to all my colleagues and friends in Edinburgh, both past and present, for making my time here so unforgettable.

ABSTRACT

An investigation into the conformational stability of the small globular protein ubiquitin (Ub) is described. Solid phase peptide synthesis has been successfully applied to the construction of both mammalian ubiquitin, as well as to a ubiquitin analogue (Ubdes-Core) lacking the protein's pronounced hydrophobic core. In each case the synthetic protein was purified to homogeneity using a combination of gel filtration, dialysis and ion exchange chromatography. Synthetic ubiquitin was found to be fully active in a ubiquitin specific *in vitro* protein conjugation assay, whereas Ubdes-Core was found to be completely inactive. The structure of these two proteins has been studied in some detail. Synthetic ubiquitin possesses an identical crystal and NMR derived solution structure to its natural counterpart. Removal of the hydrophobic core from ubiquitin results in a substantial loss in conformational stability (3.7 kcal mol^{-1}) and consequently Ubdes-Core is unable to adopt the ubiquitin native fold. This leads to the conclusion that hydrogen bonding contributions are not in themselves sufficient to stabilize the native conformation of ubiquitin.

The preparation of a 128 residue precursor protein comprising ubiquitin fused to a C-terminal extension protein (CEP52) is presented along with the CEP52 fragment alone. The compounds were prepared by solid phase peptide synthesis and purified using a combination of gel filtration, ion exchange and HPLC. Two-dimensional NMR analysis of the UbCEP52 precursor revealed, significantly, that it possesses no well defined tertiary structure. However, the protein was found to be processed *in vitro* to mature ubiquitin and CEP52. A possible novel function of ubiquitin in unfolded C-terminally fused conjugates is suggested.

CONTENTS

ABBREVIATIONS

Ac	acetyl
Acm	acetamidomethyl
Ang	angstrom
ATP	adenosine triphosphate
Boc	t-butyloxycarbonyl
Bov	bovine
Bpoc	biphenyloxycarbonyl
Bum	butoxymethyl
CEP	carboxy-terminal extention protein
CD	circular dichroism
CM	carboxymethyl
COSY	correlated spectroscopy
Δ	deletion
DCC	N,N'-dicyclohexylcarbodiimide
DCM	dichloromethane
DEAE	diethylaminoethyl
DIC	N,N'-diisopropylcarbodiimide
DMAP	4-(N,N-dimethylamino)-pyridine
DMF	N,N-dimethylformamide
DNA	deoxyribonucleic acid
DTNB	5,5'-dithiobis(*o*-nitrobenzoic acid)
DTT	dithiothreitol
E1	ubiquitin activating enzyme
E2	ubiquitin carrier protein
E3	ubiquitin-protein ligase
EDT	ethane-1,2-dithiol
EMS	ethyl methyl sulphide
FAB	fast-atom bombardment
FID	free induction decay
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast protein liquid chromatography
GdmCl	guanidinium hydrochloride
H2A	histone-2A
H2B	histone-2B
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]

HIV	human immunodeficiency virus
HOBt	1-hydroxybenzotriazole
HPLC	high performance liquid chromatography
IgG	immunoglobin G
IR	infra-red
LD-MS	laser desorption mass spectrometry
Mbh	4,4'-dimethoxybenzhydryl
MOPS	4-morpholinepropanesulphonic acid
mRNA	messenger RNA
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser and exchange spectroscopy
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PG	protecting group
Pmc	2,2,5,7,8-pentamethylchroman-6-sulphonyl
R	resin
RNA	ribonucleic acid
ROESY	rotating frame NOESY
RP	reverse phase
SDS	sodium dodecylsulphate
SPPS	solid phase peptide synthesis
TBS	tris buffered saline
TFA	trifluoroacetic acid
TFE	trifluoroethanol
TLC	thin layer chromatography
TMS	tetramethylsilane
TOCSY	total correlated spectroscopy
Tricine	N-tris(hydroxymethyl)methyl glycine
Tris	tris(hydroxymethyl)aminomethane
Trityl	triphenylmethyl
Trt	trityl
Ub	ubiquitin
UV	ultra-violet
Z	benzyloxycarbonyl

Chapter 1 - Introduction

The study of biological systems at a molecular level enables us to appreciate the complex and fascinating mechanisms fundamental to life itself. Increasingly though, our ability to make sense out of what at times appears impossibly complicated and inaccessible, relies heavily upon the use of primary metabolite probes which help to resolve this complex area of study.

In the last decade or so there have been substantial and indeed extremely innovative advances in the field of molecular biology. These have provided researchers with the tools to more efficiently study many aspects of the discipline. Recombinant DNA technology, probably the single most important development, is to a great extent responsible for the wealth of data and understanding which have been accumulated over this relatively brief period of time. However, this powerful technique does have its shortcomings. For example, by and large researchers are limited to using the 20 naturally ocurring amino acids, although it is now possible in certain cases to incorporate unnatural amino acids in moderate yields[1]. Naturally occurring amino acids bearing isotopic labels cannot be site mutated into proteins via recombinant technology, since there is no method of distinguishing the codon at the mutation site from the codons for other amino acids of that type in the protein.

Recent advances in the field of solid phase peptide chemistry have made it possible to artificially synthesise small proteins. For example, HIV-1 protease (99 amino acids) has been successfully constructed and subsequently crystallized, yielding considerable information on the mode of action of the enzyme [2]. The range of unnatural and labelled amino acids that can be incorporated into proteins via chemical synthesis is limited only by our imaginations, although is sometimes hindered by the lack of availability of the required amino acid. Chemical synthesis can in addition afford precursor proteins which normally would be processed *in vivo*. Finally, it is possible to obtain relatively large amounts of synthetic proteins, which allows for detailed structural analysis.

In our work we have chosen to study the small globular protein ubiquitin (Ub). This highly conserved polypeptide has a primary sequence of 76 amino acids and is thus of an ideal size for establishing methods of routine and efficient synthesis. The tertiary structure of this protein is well understood in both crystal [3] and solution

phases [4,5], therefore making it easier to authenticate the conformation of a synthetic ubiquitin with that of the natural material. Ubiquitin is a good model for the study of protein structure and folding, since its tertiary structure contains many of the secondary structural units commonly found within globular proteins, whilst being small enough to facilitate a comprehensive study of folding intermediates. Our aims were initially to synthesise ubiquitin in a biologically active form and thereafter to synthesise analogues of ubiquitin by the insertion of unnatural and specifically labelled amino acids which would allow investigation into the folding mechanism of the protein.

Section 1.1 Ubiquitin Biology

Ubiquitin is a highly conserved (76 amino acids, 8565 Da) protein found in all eukaryotes [6] both as the free protein and covalently attached to other proteins. It was first discovered in 1975 by Gideon Goldstein [7], who isolated it as a protein capable of inducing β-lymphocyte differentiation.

The extraordinary conservation of ubiquitin throughout eukaryotes (see figure 1.1) points to a fundamental involvement in cellular functions and indeed since its discovery it has been associated with an astonishing array of biological processes, many of which will subsequently be discussed.

	5			10			15			20			25					
Animal	M	Q	I	F	V	K	T	L	T	G	K	T	I	T	L	E	V	E
Oat																S		D
Yeast																S		D

	P	S	D	T	I	E	N	V	K	A	K							
Animal																		
Oat																		
Yeast							S											

	30			35			40			45			50			55		
Animal	I	Q	D	K	E	G	I	P	P	D	Q	Q	R	L	I	F	A	G
Oat																		
Yeast																		

	K	Q	L	E	D	G	R	T	L	S	D							
Animal																		
Oat										A								
Yeast																		

	60			65			70											
Animal	Y	N	I	Q	K	E	S	T	L	H	L	V	L	R	L	R	G	G
Oat																		
Yeast																		

Figure 1.1 Primary Sequence of Ubiquitin in Plant and Animal.

1.1.1 The Ubiquitin Gene

There have been a number of ubiquitin encoding genes isolated from a variety of organisms. Analysis of these data reveals that ubiquitin genes can be divided into two discrete categories, namely those which encode polyubiquitin precursors and those which encode ubiquitin fused to C-terminal extension proteins (CEPs) bearing no resemblance to ubiquitin.

A series of genes has been identified in man [8], chicken [9], yeast [10], xenopus laevis[11] and plant cells [12], all of which encode a sequential pattern of ubiquitin units connected in a head to tail manner with no internal spacer elements. In each case the C-terminus of the gene codes for a single extra amino acid followed by a stop codon. The number of ubiquitin repeats varies according to source as indeed does the type of amino acid at the C-terminal (see figure 1.2).

$$\underset{\text{Met}}{1}\text{ - }\underset{\text{Gly}}{76}\text{ - (}\underset{\text{Met}}{1}\text{ - }\underset{\text{Gly}}{76}\text{)}_n\text{ - }\underset{\text{Met}}{1}\text{ - }\underset{\text{Gly}}{76}\text{ - X}$$

X = Val - human, n = 9 or 3
X = Asn - yeast, n = 5
X = Tyr - chicken, n = 5

Figure 1.2 Polyubiquitin Genes

Proteolytic cleavage at the Gly-Met and Gly-X junctions would be required to liberate mature ubiquitin from the precursor. The ubiquitin carboxy-terminal amidases responsible for these hydrolyses have been isolated from bovine calf thymus [13], but are yet to be fully characterized. Exactly how quickly after translation this processing occurs is as yet unknown, however the single amino acid extension presumably acts to prevent the precursor undergoing reactions at its C-terminus (part of mature ubiquitin's mode of action). This would imply that the unprocessed material was relatively long lived. Polyubiquitin precursors may well exist as a form of cytoplasmic ubiquitin store which could be rapidly processed when required.

The second category of ubiquitin gene which has been identified encodes for ubiquitin fused through the C-terminus to a highly basic extension protein (so called CEPs). These extension segments, isolated from human [14,15], dictyostelium [16] and yeast [17] genomes, can be either 52 or 76-80 amino acids in length and indeed are highly conserved (80%) between these organisms. Unlike the polyubiquitin case, the lifetime of these precursors is thought to be relatively short. In fact, it has been proposed that the processing may even occur co-translationally[18]. The biology of UbCEPs will be discussed in some detail in a later section.

1.1.2 Ubiquitin and Chromatin Structure

The earliest role of ubiquitin to be identified was an involvement in chromatin structure. Sequencing of the nuclear protein, A24, by Busch and Goldnopf had revealed the existence of a branched structure in which ubiquitin was a component[19]. Subsequent work by Hunt and Dayhoff established that this branched structure involved the conjugation of the C-terminal glycine of ubiquitin to the ε-amino group of lysine-119 of histone-2A [20]. It has since been shown that ubiquitin forms conjugates with histone-2B [21] (uH2B) and with all subtypes of histone-2A [22,23] (uH2A , see figure 1.3).

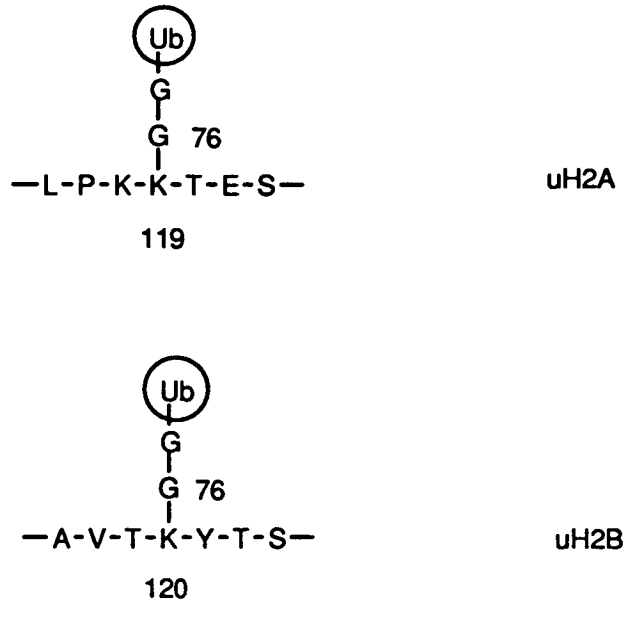


Figure 1.3 Branching Region in uH2A & uH2B

Histones are nuclear proteins which contain a large proportion of positively charged lysine and arginine residues, obviously allowing the protein to complex with the negatively charged phosphate backbone of DNA. The four types of histone involved in chromatin structure form an octamerous aggregation around which the DNA diploid is wound forming a negatively coiled superhelix. Up to 10% of histone-2A[24] and 1.5% of histone-2B [21] can be ubiquitinated within the resulting nucleosome.

Histones 2A and 2B have been shown to be polyubiquitinated (pu) in trout livers[25], and the structure of puH2A has been studied in detail using residue specific proteolysis [26]. Peptide mapping of the resulting digest points to individual ubiquitin molecules being joined to one another via isopeptide bonds between the C-terminal glycine of one ubiquitin and an ε-amino lysine group of the next ubiquitin. Ultimately the polyubiquitin unit is again attached to the histone through lysine-119.

Ubiquitination of histones was initially thought to have a direct role on gene regulation. Early results indicated that actively transcribing avian erythrocytes had a much higher level of uH2A than the corresponding transcriptionally inactive chicken erythrocytes [27]. Other studies have similarly suggested that ubiquitination of histones is in some way associated with increased transcription within the nucleus [28,29,30]. However, it is significant that inhibition of transcription in HeLa cells is not accompanied by a short-term fall in uH2A levels [31]. It seems more likely that the ubiquitination of histones is implicated in overall chromatin structure, and that its apparent involvement in transcription is actually a consequence of the association of DNA binding proteins with the nucleosome, leading to a change in that structure. Levels of uH2A do not remain constant throughout the cell cycle. In fact, there are conspicuous fluctuations which tend to reinforce the proposed involvement in chromatin structure. Studies have shown [32] that the amount of uH2A in the nucleus falls to a minimum just before metaphase, thereafter returning to normal levels by the G1 stage in the cell cycle (see figure 1.4). These results imply a role in cell mitosis and indeed it has been suggested that de-ubiquitination of histones, by a specific isoamidase[13], is the trigger for chromosomal condensation.

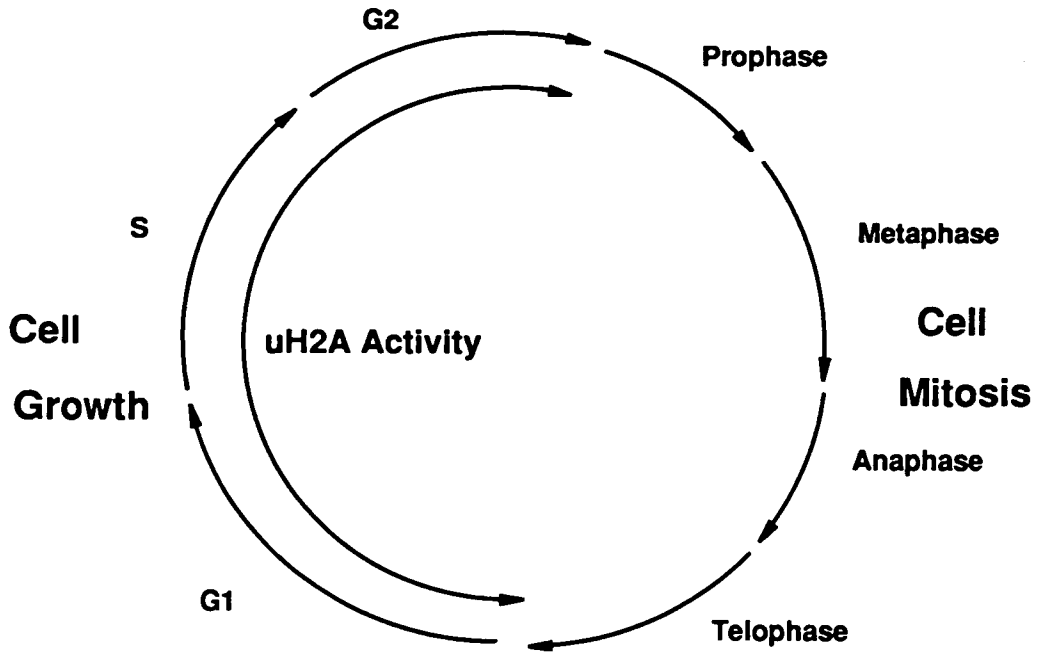


Figure 1.4 Variation of uH2A levels during the Cell Cycle

Recently it has been reported [33] that cyclin, a regulatory protein whose proteolysis precedes exit from cell mitosis, is ubiquitinated before being degraded via the ubiquitin degradation pathway (see section 1.1.5). It may therefore be the case that ubiquitin is involved in both the entrance to and exit from cell mitosis. As yet it is unknown what, if any, conformational changes occur upon ubiquitination of a histone, and it is also unclear why some histones are polyubiquitinated and not others. However, it is clear that the involvement of ubiquitin in chromatin is an important variable in the cell cycle, and therefore of some significance.

1.1.3 Ubiquitin and Cell Surface Structures

The functions of ubiquitin are not restricted to the intracellular worlds of the cytoplasm and nucleus but are known to extend to various cell surface structures. The first identified example of such a system, the lymphocyte homing receptor, was found to be ubiquitinated by Weissman and co-workers [34,35]. Lymphatic organs such as the thymus, lymph nodes and spleen are constantly fuelled with mature lymphocytes in a process known as homing. Entry of the correct type of lymphocyte into the organ, through blood vessels called high endothelial venules (HEV 's), is controlled by specific lymphocyte homing receptors. Weissman and co-workers used a monoclonal antibody (Mel-14) in order to study the murine

receptor. This antibody, specific for peripheral lymph nodes, was used to locate the gene coding for the homing receptor on these organs. Cloning of this gene was followed by purification of the Mel-14 reactive antigen by immunoprecipitation. Analysis of this glycoprotein revealed the existence of two N-termini, one of which corresponded to that of ubiquitin. Subsequent analysis with monoclonal antibodies raised to ubiquitin, have confirmed that the 90 kDa cell surface protein is ubiquitinated. It is interesting that while Mel-14 has its epitope in the C-terminal region of ubiquitin, it appears not to recognize this protein when it is not conjugated to the rest of the receptor. This implies that ubiquitin undergoes a conformational change upon conjugation. Little is known about the number of ubiquitin molecules covalently attached to the receptor, however, ubiquitination is thought to occur fairly early on in the receptors biosynthesis [36], and in a manner similar to that of histones. It seems equally unclear why these receptors are ubiquitinated, although it is possible that this has implications for the structure of the receptor.

Two receptors involved in cell replication are believed to be ubiquitinated. Sequencing of the receptors specific for platelet-derived growth factor [37] and somatotropin [38] has, in each case, revealed the existence of two N-terminal sequences, one of which was identified as that of ubiquitin. It is not known whether ubiquitin is attached to the extracellular domain or the intracellular domain of these receptors. The role played by ubiquitin in these membrane bound proteins is equally unclear, however, two different functions have been proposed. Firstly, ubiquitination may be the signal for proteolysis of these relatively short-lived receptors (see later for ubiquitin dependant proteolysis). Clearly the proliferation of these receptors would have to be carefully regulated in order to prevent malignant cellular growth. It has also been proposed that receptor ubiquitination/de-ubiquitination is involved in signal transduction in a manner analogous to that of kinase activity.

A novel autocrine growth factor, involved in myeloma and lymphoid cell growth, may be a ubiquitin-protein conjugate [39]. Initial studies have shown that the N-terminal sequence of leukemia derived growth factor corresponds to that of ubiquitin. Although virtually nothing is known about this growth factor, the possibility that its membrane bound receptor also has a ubiquitin component does

reinforce the suggestion that ubiquitin has a regulatory role within the autocrine system.

1.1.4 Ubiquitin and Neural and Muscular Degeneration

Alzheimer's disease, the principle cause of dementia in aged humans, is characterized by the progressive formation of cytoplasmic neurofibrillary tangles in brain tissue [40].

A monoclonal antibody specific for paired helical filaments, the major constituent of these neuronal fibres, was found to recognize ubiquitin [41]. Subsequent studies involving proteolytic digestion followed by sequencing of the resulting peptides, confirmed that paired helical filaments were ubiquitinated. This initial observation was followed by the discovery that ubiquitin specific antibodies recognized the neuronal fibres in a variety of other neurodegenerative diseases such as Down's syndrome and Creutzfeldt-Jakob disease [42].

The insolubility of neuronal fibres has resulted in their pathology being poorly understood and consequently one can only speculate as to why they are ubiquitinated. Ubiquitin, however, is not evenly distributed throughout a neuronal fibre, instead being localised within subpopulations. This suggests it is, by in large, not involved in the formation of the inclusion body. Instead, it has been proposed that it is involved in the cellular proteolytic response to the formation of these cytoplasmic tangles [41].

The involvement of ubiquitin in muscular degeneration is far better understood. Deterioration of muscle tissue, frequently a non-pathological process, can in some cases be accompanied by an increase in polyubiquitin gene translation [43]. This is thought to be because degeneration associated proteolysis occurs via the ubiquitin pathway.

1.1.5 Ubiquitin and ATP-Dependant Proteolysis

The ATP-dependant proteolytic system is easily the most vibrant and indeed best understood area of ubiquitin research. There has been quite a wealth of publications on this ubiquitin dependant system and recent reviews cover the area in some detail [44,45].

Turnover of cytoplasmic proteins is extremely important to the continuing vitality of a cell, and it is in this area that ubiquitin appears to perform its most fundamental role. Early investigations found that ion exchange chromatography could resolve murine reticulocyte lysates into two fractions (I & II) and that separation of these resulted in the loss of the normal proteolytic activity associated with the unresolved lysate [46]. Ubiquitin was subsequently shown to be the active component in fraction I [47] and analysis of fraction II has revealed the existence of a multi-enzymic conjugation system which mediates the attachment of ubiquitin to proteins destined for degradation by protease [48]. As with the nuclear protein uH2A, the C-terminal of ubiquitin forms an isopeptide link with the ε - group of a lysine residue and ultimately a target protein may be multi-ubiquitinated [49].

1.1.5(i) Mechanism of Ubiquitin Conjugation

The covalent attachment of ubiquitin to a substrate protein utilises a specific ligase system of three distinct enzymes, which together activate the C-terminus of ubiquitin in a four step process (see figure 1.5).

1. In the first step, the C-terminus of ubiquitin is adenylated [50]. Furthermore, studies involving radiolabelled pyrophosphate, released during the reaction, indicate this process occurs in the presence of the so called ubiquitin activating enzyme (E1).

2. The ubiquitin C-terminal adenylate then reacts with a cysteine thiol group on the E1 enzyme, resulting in the formation of a thioester [51].

3. A transthiolation step involving a second enzyme, ubiquitin carrier protein (E2), gives rise to a new thioester linkage capable, in some cases, of transporting ubiquitin to a substrate protein [52].

4. Actual conjugation of ubiquitin to a protein destined for degradation usually requires the presence of a third class of enzyme. Ubiquitin-protein ligase (E3) is thought to catalyse the formation of the final isopeptide bond [52].

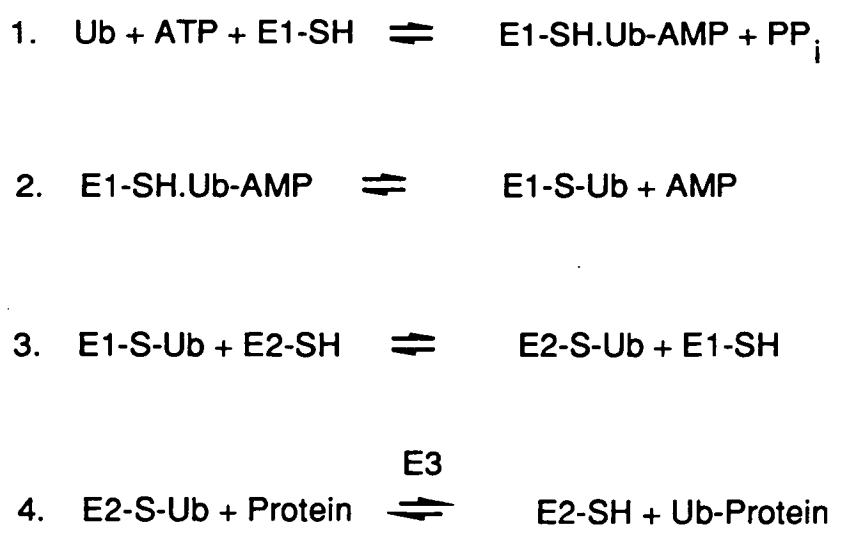


Figure 1.5 Chain of Events in Ubiquitin Protein Conjugation

1.1.5(ii) Enzymology of the Ligase System

Several forms of E1, E2 and E3 have been isolated from a variety of organisms, and this has led, along with analysis of the corresponding gene sequences, to information being obtained on their structures.

A ubiquitin affinity column was used to isolate E1 from murine reticulocyte fraction II [53]. This enzyme, a homodimer with a subunit of 110 kDa, appears to be the only form of E1 in murine reticulocytes. Two apparently different forms of E1 have been been identified in humans [54,55], however, both are thought to act as homodimers. The inability of many mammalian cells to tolerate specific mutations in E1 suggests the existence of only one form of the enzyme in these organisms[54,56]. A similar argument undeniably points to there being a single E1 gene in yeast [57]. Multiple forms of E1 have been isolated from wheat germ [58], again using a ubiquitin affinity column. Analysis of this heterogeneous mixture of E1 enzymes by SDS and Native-PAGE, indicates they exist as monomers, all having a molecular weight of around 120 kDa.

The isolation of several E2-encoding genes from yeast may help to explain the extraordinary selectivity of the ubiquitin-dependant proteolytic system [59]. These E2 enzymes differ in their molecular weight and perhaps significantly in their requirement for the E3 enzyme in transferring ubiquitin to a substrate protein [60]. Examination of their gene sequences reveals the existence of a conserved core

region of about 150 amino acids (containing the active cysteine) around which there may or may not be a C-terminal extension sequence [59]. Three of these genes, designated UBC4, UBC5 and UBC7, encode E2 enzymes which require an E3 enzyme to effect final ubiquitin-protein conjugation. Interestingly, these three enzymes consist only of the central conserved region. However, the UBC1, UBC2 (RAD6) and UBC6 genes, all of which have various C-terminal extensions, encode E2s able to operate independently of the E3 enzyme [59]. In these cases, it is not unreasonable to suggest that the extensions play a "E3-like" role in recognizing a substrate protein, indeed this proposal is supported by the fact that E2s having highly acidic C-terminal extensions can readily conjugate ubiquitin to extremely basic histones [61].

The UBC2 (RAD6) enzyme appears to have a DNA repair function, this is most probably connected with its ability to conjugate ubiquitin to histones [61]. Ubiquitination of a histone complexed to damaged DNA may well mark the site for binding of repair enzymes or, alternatively, lead to exposure of the damaged DNA by targeting the histone for degradation [62]. Genetically removing the C-terminal tail from RAD6 does not appear to completely eliminate this DNA repair function[63], perhaps suggesting there are substrate recognition features within the remaining core region of the enzyme.

Two E3 enzymes have, to date, been isolated from rabbit reticulocytes [64,65] and a yeast gene (UBR1) has recently been identified and cloned [66]. The two murine E3s, known as E3α and E3β, have different binding affinities towards substrate proteins (see N-end rule 1.1.5(v)). Proteins with bulky hydrophobic or basic residues at their N-termini are recognized by E3α, whereas E3β appears to recognize small uncharged amino acids [67]. At the present time it is not known whether a class of E3 enzymes recognizing internal sequences exist.

The possibility of proteins being incorrectly or unnecessarily ubiquitinated leads to a requirement for de-ubiquitination enzymes which would play a role in maintaining the integrity of the proteolytic system. However, such a requirement will not always be due to errors in the conjugation process; we have already seen that the liberation of mature ubiquitin from polyubiquitin and UbCEP precursors is carried out by specifically tailored hydrolases [13,18]. Clearly there would also be a demand for enzymes with the ability to regenerate ubiquitin from conjugates

arising from the reaction of E1-Ub or E2-Ub with any one of the numerous cellular nucleophiles such as glutathione. De-ubiquitination of isopeptide linked ubiquitin-protein conjugates is carried out by a third class of enzyme, of which five have have been identified [13,68]. These isopeptidases, all from mammalian cells, appear to have different specificities relating to the particular cell type to which they belong.

1.1.5(iii) ATP-Dependant Protease of the Ubiquitin System

While a great deal is known about how ubiquitin becomes attached to a substrate protein, relatively little information is available on the subsequent mode of degradation of that protein.

Initial observations noted that the digestion required both ATP and Mg^{2+} ions [69]. Subsequent use of ubiquitin-lysozyme conjugates as substrates, enabled Rechsteiner and co-workers to identify the protease that degrades proteins conjugated to ubiquitin in a ATP dependant manner [70]. This protease, isolated from rabbit reticulocytes, was found to have an extremely high molecular weight (26S, 1500 kDa) and indeed other protease of comparable size have since been detected in other organisms [71]. Ganoth and co-workers found that reticulocyte cells starved of ATP contained three different factors, all of which were required for degradation of proteins tagged with ubiquitin [72]. These three components (each of 250-650 kDa) apparently assemble to form the 26S protease in the presence of ATP and Mg^{2+} [72].

One of these has been identified as the 20S proteasome [73], a multi-catalytic cylindrical protease complex which was previously thought to be distinct from the 26S protease. Immunoprecipitation of the proteasome with specific antibodies was found to inhibit the 26S protease, implying the proteasome is the catalytic centre of the complex [74]. The second, 250 kDa, factor of 26S is thought to inhibit some of the catalytic functions of the proteasome, thereby modifying its activity to suit the ubiquitin dependant system [75]. As yet, the role played by the third, 600 kDa, component of 26S is unknown, although there has been speculation about a possible proof-reading function. Alternatively, it could act as a structural template, allowing the other two components to be correctly oriented within the overall complex.

In a recent paper, Hershko and co-workers reported the isolation of a 100 kDa monomeric ubiquitin binding protein abundant in erythrocytes and reticulocytes [76]. This enzyme was purified from fraction II by a combination of ubiquitin-affinity chromatography and subsequent gel filtration chromatography. The enzyme apparently converts high molecular weight polyubiquitin-protein conjugates into lower mass forms with the liberation of mature ubiquitin, but importantly not of free protein substrate. Surprisingly, this isopeptidase appears to actually stimulate the degradation of substrate proteins by the 26S protease complex [76]. This implies that some polyubiquitin-protein conjugates are simply too large for the 26S protease to deal with and that this isopeptidase effectively tailors them into a convenient size. Furthermore, addition of the isopeptidase to an *in vitro* system in which polyubiquitin-protein conjugates had been pre-incubated with the 26S complex, resulted in high levels of free ubiquitin being formed. Therefore, the isopeptidase appears to regenerate the ubiquitin from the polyubiquitin remains of the degradation process. Figure 1.6 illustrates a possible order of events in the degradation process.

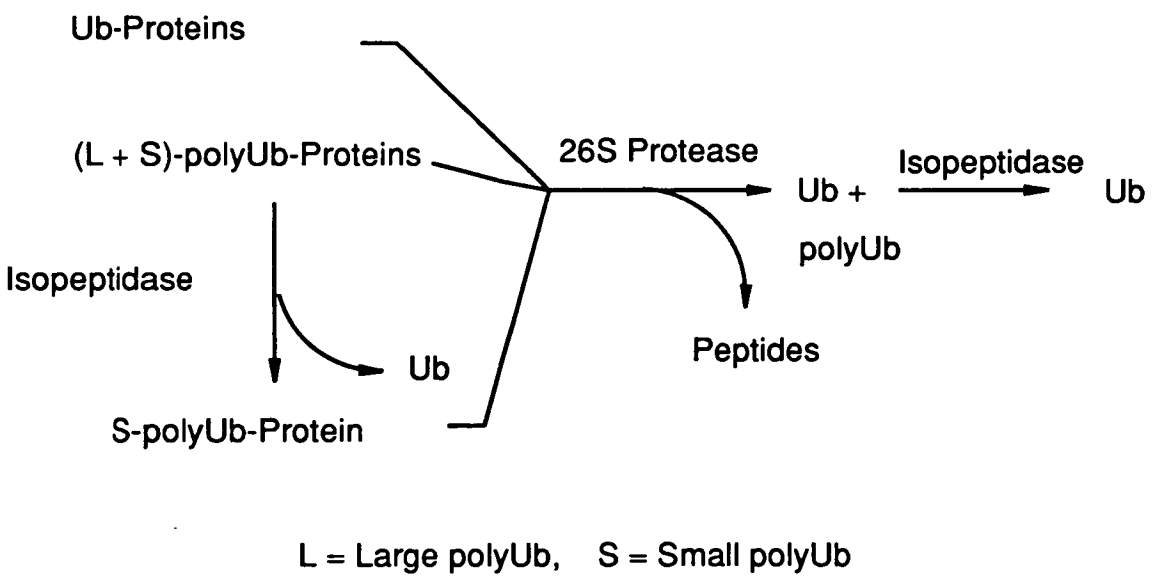


Figure 1.6 Proposed Degradation Pathway

1.1.5(iv) Turnover of Proteins via the Ubiquitin Pathway

Attempting to organize cellular proteins into groups according to their half-lives is a somewhat capricious affair, however, most proteins fall into one of the following three general classes.

1. Short Lived Proteins. These are endogenous proteins with a high turnover rate.

2. Long Lived Proteins. The vast majority of cytoplasmic and nuclear proteins have a relatively slow rate of turnover.

3. Abnormal Proteins. This class of proteins would include exogenous proteins, as well as damaged or mutated endogenous proteins. Turnover of these would necessarily be the highest of all.

Proteolysis of short lived intracellular proteins appears to principally occur via the ubiquitin pathway, whereas, long lived proteins are thought to undergo lysosomal degradation [77]. Radiolabelling studies have shown that many abnormal exogenous proteins also undergo degradation via the ubiquitin pathway [78]. Autoradiography of reticulocyte cell contents clearly shows that these proteins are both mono- and multi-ubiquitinated.

Whilst the ubiquitin pathway is usually associated with cytosolic protein degradation, there is some controversy over whether it also operates within lysosomes. Docherty and co-workers found ubiquitin-protein conjugate levels to be greatly enhanced within lysosomes relative to cytosolic levels, clearly suggesting a lysosome-ubiquitin affiliation [79]. However, other studies indicate that ubiquitin-protein uptake is non-selective and that their accumulation is a result of lysosomes not possessing the required 26S protease to degrade them [80]. In a further study, Gropper and co-workers have shown that an E1 enzyme is required for stress-induced lysosomal degradation of long lived proteins [81]. This group ultimately propose that ubiquitin activation is somehow required for the formation of autophagic vacuoles, the vessels through which proteins enter lysosomes.

1.1.5(v) Importance of the N-terminus of Substrate Proteins

It is estimated that around 80% of cellular proteins are N-terminally acetylated [82]. Of the remaining 20%, the nature of the free N^{α}-amino residue is crucially important in determining these proteins rate of turnover via the ubiquitin pathway (see N-end rule). Clearly though, most proteins must have a signal downstream of their blocked N-termini which controls how quickly they are tagged with ubiquitin

for degradation. As already mentioned, this "internal signal" may be recognized by an as yet unknown class of E3 enzyme.

N^{α}-acetylated proteins are not degraded in reticulocyte lysate fraction II supplemented with ubiquitin, however, these proteins are degraded in whole lysates in a ubiquitin dependant manner [83]. Ciechanover and co-workers have recently isolated a protein from reticulocyte fraction I which appears to be the active component required for the ubiquitin-dependant degradation of N^{α}-acetylated proteins [84]. This enzyme, designated Factor H, is a homodimer of subunit size 46 kDa, and was shown to be required for the degradation of ubiquitinated histone-2A, an N^{α}-acetylated protein. Interestingly, this group were unable to form ubiquitin conjugates to other N^{α}-acetylated proteins, including actin and α-crystallin, this was thought to be because the *in vitro* conjugation system contained only E3α and E3β, thereby implying a third form of E3 is generally required for this type of conjugation. While little is known about the mode of action of Factor H, it presumably must in some way interact with conjugates before proteolysis can occur.

1.1.5(vi) The N-End Rule

A remarkably straightforward rule governing the susceptibility of free N^{α}-amino proteins to degradation has been discovered by researchers performing experiments utilising site directed mutagenesis [85]. All twenty amino acids have been systematically site mutated at the N-terminal of the long lived bacterial protein, β-galactosidase, and the turnover of the resulting twenty variants investigated in both yeast [85] and more recently reticulocytes [86]. As with most proteins the first amino acid in wild-type β-galactosidase is methionine, whose codon (AUG) also happens to be the start codon for gene transcription. Mutation of N-terminal methionine can therefore lead to certain problems. However, Varashavsky overcame these by constructing a plasmid encoding ubiquitin fused to the N-terminus of β-galactosidase, which results in the 19 mutations being located at the junction [85]. Following expression, the fusion protein is rapidly de-ubiquitinated, thereby exposing mature β-galactosidase bearing the required N-terminal mutation.

Studies revealed that certain amino acids are destabilizing when located at the N-terminal of the protein, whereas others apparently have the opposite effect. The turnover rate of the former class of proteins is markedly higher than the later (see

table 1.1). This seemingly simple relationship between the cellular lifetime of a protein and the identity of its N-terminal residue is known as the N-end rule. Indeed, the order of susceptibility is similar in both reticulocytes and yeast, with the differences being ascribed to the presence of E3β in the former but not the later.

N-TERMINAL AMINO ACID INCORPORATED INTO β–GALACTOSIDASE		CELLULAR HALFLIFE	CELLULAR HALFLIFE
Primary Destabilizing Residues		**Yeast**	**Reticuloctyes**
Type I	Arginine	2 minutes	1 hour
	Lysine	3 minutes	1.3 hours
	Histidine	10 minutes	3.5 hours
Type II	Phenylalanine	3 minutes	1.1 minutes
	Leucine	3 minutes	5.5 hours
	Tryptophan	3 minutes	2.8 hours
	Tyrosine	10 minutes	2.8 hours
	Isoleucine	30 minutes	20 hours
Type III	Alanine	>20 hours	4.4 hours
	Serine	>20 hours	1.9 hours
	Threonine	>20 hours	7.2 hours
Secondary Destabilizing Residues			
	Aspartic acid	3 minutes	1.1 hours
	Glutamic acid	30 minutes	1 hour
	Cysteine	>20 hours	1.2 hours
Tertiary Destabilizing Residues			
	Asparagine	3 minutes	1.4 hours
	Glutamine	10 minutes	0.8 hours

Table 1.1 Comparison of Destabilizing Residues in Yeast and Reticulocytes

The destabilizing effect of certain N-terminal amino acids can be neatly explained by the affinity E3 enzymes have towards them. Yeast and reticulocytes both contain E3α, an enzyme which apparently recognizes basic type I residues, as well as bulky hydrophobic type II amino acids. Evidence for E3α having two separate binding sites comes from the observation that dipeptides with type I or II residues at their

N-termini are full antagonists of the enzyme [87]. Reticulocytes, which also contain E3β, have a third class of destabilizing amino acid. These small uncharged residues are, however, not destabilizing in yeast, implying that type III residues bind to E3β.

The above explanation cannot directly account for the destabilizing effects of Asp, Glu, Asn and Gln residues, none of which are recognized by E3α or E3β. However, studies have shown that proteins with acidic amino acids at their N-termini can be post-translationally modified, resulting in an arginine residue being added onto their N-termini [89]. Such arginylation, resulting in recognition by E3α, is thought to be catalysed by a multi-enzymic system comprising of arginyl-tRNA-protein transferase and arginyl-tRNA synthetase [89]. Aspartic acid and glutamic acid are therefore known as secondary destabilizing residues, whereas, asparagine and glutamine, which initially have to hydrolysed to the corresponding acids, are referred to as tertiary destabilizing residues.

1.1.5(vii) Multi-Ubiquitination of Protein Substrates

Clearly not all ubiquitinated proteins are targeted for proteolysis, for example, nuclear histone proteins are for the most part relatively stable when ubiquitinated. The possibility that multi-ubiquitination of a protein can somehow convey the necessary information for or against degradation, has a certain elegant attraction.

Addition of exogenous lysozyme to reticulocytes results in ubiquitin-lysozyme conjugates in excess of a 100 kDa being formed [69,90,91]. Lysozyme, with seven free amino groups, would be expected to have a maximum weight of 74 kDa if mono-ubiquitination were to occur at every site. The extra weight can be accounted for by the formation of multi-ubiquitin chains similar to those found in histone-2A[25]. Chemical analysis of these conjugates subsequently confirmed this, as well as suggesting adjacent ubiquitin molecules in the chain are linked through isopeptide bonds [91,69].

A similar situation is found in β-galactosidase and, furthermore, peptide mapping of this ubiquitin-protein conjugate indicated that the afore mentioned isopeptide bonds occur between the C-terminal of one molecule and the ε-amino group of Lys-48 in the next [91]. A mutant ubiquitin in which Lys-48 had been changed to an arginine residue (unable to form isopeptide bonds) was found to yield only mono-ubiquitinated β-galactosidase *in vitro* [91]. In addition, this ubiquitin mutant could

not sustain the *in vitro* proteolysis of β-galactosidase by the 26S protease complex, thus, implying multi-ubiquitination is necessary for protein degradation.

It seems likely that multi-ubiquitination is, in most cases, a requirement for substrate degradation, however, exactly why this is so is still under dispute. One can envisage a whole range of branched multi-ubiquitin "flags", each one signaling a specific turnover rate for the conjugated protein. Alternatively, multi-ubiquitination may simply be a way of amplifying a considerably weaker mono-ubiquitin signal.

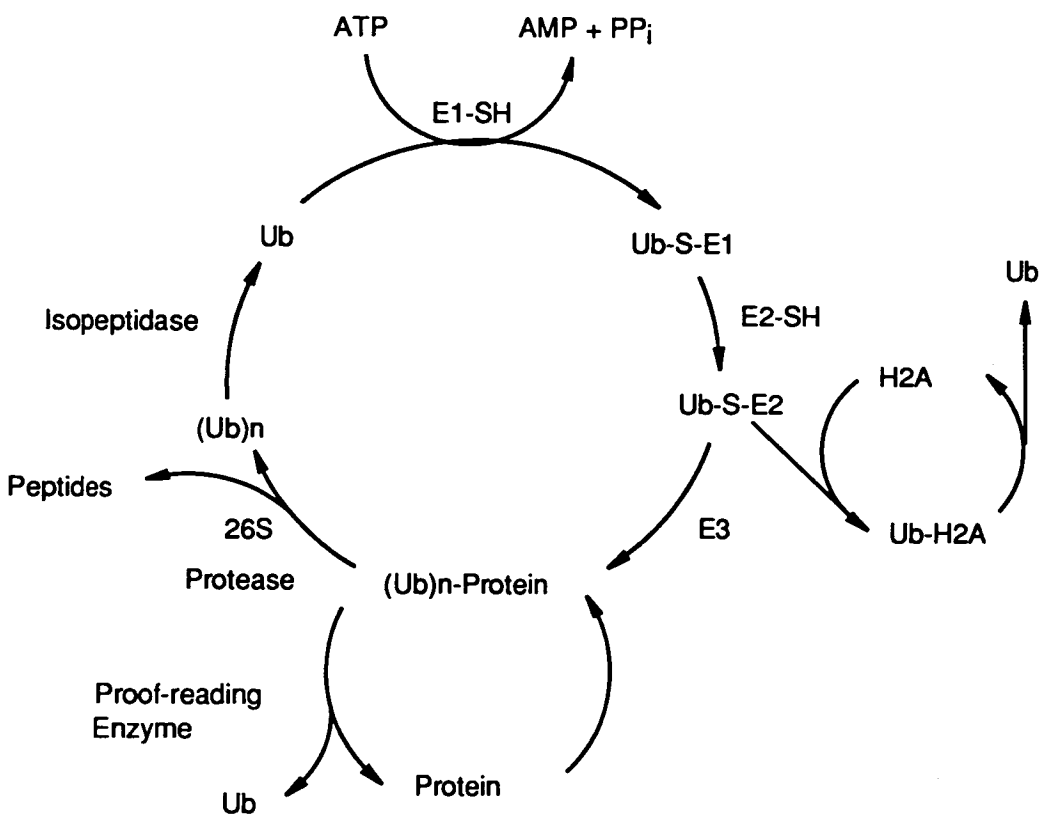


Figure 1.7 The Ubiquitin Pathway

1.1.5(viii) Overall Picture

In summary, we have seen how ubiquitin becomes attached to a substrate protein through activation of its C-terminal by a specific set of enzymes (E1, E2 and E3).

The existence of a group of proof-reading enzymes which prevent needless and erroneous ubiquitination has also been described. A protein's rate of turnover, by the 26S protease specific for the ubiquitin pathway, is dependant upon the number of ubiquitin molecules conjugated to it, which in turn may be dependant upon specific variables in the protein's primary structure. These include the N-terminal amino acid or, in the case of N-terminally acetylated proteins, possible internal signal sequences. Figure 1.7 attempts to illustrate the events in the ubiquitin pathway.

1.1.6 Ubiquitin and The Heat Shock Response

Exposure of cells to temperatures above the physiological norm, results in the synthesis of a specific set of proteins. These heat shock proteins, the cells response to environmental stress, are also induced by exposure to toxic chemicals, starvation and various other metabolic challenges. In whatever way it manifests itself, cellular stress would be expected to bring about the build up of damaged or denatured proteins.

Early evidence for the involvement of ubiquitin in the heat shock response came from work carried out on the mouse carcinoma cell-line, ts85 [92,93]. These mutant cells contain a thermolabile ubiquitin activating enzyme (E1) which is inactivated at 39.5 ^{o}C, the non-permissive temperature. This observation implied that at the elevated temperature, ubiquitination of proteins is repressed and that this in some way triggers the synthesis of heat shock proteins. However, this interpretation is undermined by a later study in which ^{125}I-ubiquitin was shown, using autoradiography, to conjugate very small amounts of substrate proteins at the elevated temperature [94]. Nevertheless, degradation of long lived cellular proteins, which normally increases under heat shock [95], is greatly obstructed in ts85 cells, suggesting that proteolysis of denatured proteins is ubiquitin dependant. That heat shock induces ubiquitin synthesis in ts85 cells [92,93] and in chicken embryo fibroblasts [96], surely affirms the above hypothesis that degradation of the abnormal proteins produced during heat shock goes via the ubiquitin pathway.

It has been estimated that the increased rate of protein ubiquitination, which occurs during heat shock, results in the cellular ubiquitin concentration dropping to around 25% of the normal level [97]. Activation of ubiquitin genes is required to prevent cells becoming starved of the protein during times of stress. Finlay and co-workers

found that in yeast, the ubiquitin UBI4 gene becomes activated during heat shock[98]. This polyubiquitin gene is virtually dormant during normal physiological conditions throughout which yeast cells rely on three other ubiquitin genes (UBI1-3) to maintain regular pools. Two ubiquitin conjugating enzymes (E2s, UBC4 and UBC5) also appear to be induced by heat shock [99].

Although other heat shock proteins have been identified, ubiquitin is the only one known to be involved in cytoplasmic protein degradation. Three other heat shock proteins, of subunit sizes 89 kDa, 20 kDa and 24 kDa , have been isolated from chicken fibroblasts. Moreover, proteins of similar size and bearing high levels of homology have been found induced in heat shocked drosophila [100]. Since none of these has any known connection with cellular proteolysis, one can easily speculate on the possible interactions they may have with misfolded proteins, resulting in their renaturation and hence rescue from the imminent destruction afforded by the ubiquitin system.

1.1.7 Ubiquitin and Ribosome Biogenesis

A family of genes encoding ubiquitin fused to a carboxy-terminal extension sequence has been identified in the human [14,15], dictyostelium [16] and yeast [17] genomes (see figure 1.8).

These highly basic carboxyl extension proteins (CEPs) are either 52 or 76-80 amino acids long and are always synthesised in the accompaniment of ubiquitin. Moreover, both classes of CEP undergo rapid de-ubiquitination either during or very soon after translation [18,101]. Within each class there is a high degree of homology between organisms, and furthermore, every CEP contains a putative "zinc finger" DNA binding motif of the $(Cys)_4$ variety (see ref 102 for a review zinc fingers).

Speculation over the cellular fate of CEPs was resolved when Redman and Rechsteiner found that the 80 amino acid extension protein (CEP80) is ultimately located within a ribosomal subunit [103]. Antibodies specific for CEP80 reacted with the previously characterized ribosomal protein, S27a, which is a component of the small 40S ribosomal subunit. Further evidence for CEP80 being S27a came from amino acid composition and two-dimensional electrophoresis. The smaller CEP52 subclass has since been identified as a component of the larger 60S ribosomal

subunit [104,105]. Finley and co-workers have generated yeast mutants bearing deletions of the UbCEP encoding genes, UBI 1-3 [104]. This group found that those mutants lacking UbCEP52 (ubi1Δ and ubi2Δ) were deficient in the ribosomal 60S subunit, whereas the UbCEP80 deletion mutant (ubi3Δ) had lower than normal 40S subunits. In the same study, it was found that the removal of the ubiquitin encoding element from the UBI3 gene, leaving only the CEP80 component, resulted in lower levels of the 40S ribosomal subunit being produced.

```
          1                   10                  20
Ubiquitin - I-I-E-P-S-L-R-Q-L-A-Q-K-Y-N-C-D-K-M-I-C-R-K-C-Y-A-R-L-H-P-
Ubiquitin -A-L-L-R-L-L-L-S-Y-T-T-P-K-K-N-K-H-K-R-K-K-V-K-L-A-V-L-K-Y-

          30                  40                  50
- R-A-V-N-C-R-K-K-K-C-G-H-T-N-N-L-R-P-K-K-K-V-K-CO2H
- Y-K-V-D-E-N-G-K-I-S-R-L-R-R-E-C-P-S-D-E-C-G-A-G-V-F-M-A-

          60                  70
-S-H-F-D-R-H-Y-C-G-K-C-C-L-T-Y-C-F-N-K-P-E-D-K-CO2H
```

Figure 1.8 Primary Sequence of Human UbCEP52 & UbCEP80

Interestingly, the yeast genes, UBI1-3, all of which encode UbCEPs, are most active during times of cell growth. This is obviously a time when protein synthesis, and hence ribosome biogenesis, is at a maximum level. However, the remaining ubiquitin encoding yeast gene, pUBI4, is not at all active during normal cell growth, instead being maximally expressed in times of cellular stress. This increase in polyubiquitin precursor synthesis is accompanied by the shut-down of UbCEP production, presumably because during cellular stress ribosomal protein synthesis is naturally at a minimum and protein degradation is a higher priority.

Exactly why CEPs are synthesised N-terminally fused to ubiquitin is a question which, as yet, has not been answered unequivocally. It has been suggested that ubiquitin increases the incorporation of CEPs into ribosomes in a way similar to a molecular chaperon type protein [104]. In this model ubiquitin would not only aid the actual biogenesis step, but would also act as a guardian against degradation right from the moment the CEP is translated. The fusion of ubiquitin to the N-

terminal of the CEP may block any N-end rule degradation signal it might have and hence protect it from ubiquitin mediated proteolysis. Recent evidence, however, suggests that de-ubiquitination of UbCEPs occurs almost co-translationally, thus making the alleged chaperon function of ubiquitin very difficult to justify [18].

Given that deletion of the ubiquitin encoding element from the yeast UBI3 gene results in lower levels of the remaining CEP component reaching its final ribosomal destination [104], it has been proposed that ubiquitin is required for enhanced translation of the extension protein [18]. This theory is given some credibility in the light of the observation that N-terminally fused ubiquitin substantially increases the expression of heterologous gene products in E. coli and yeast [106-108]. For example, the expression in E. coli of a plasmid encoding ubiquitin N-terminally fused to human poly(ADP-ribose) polymerase results in this protein being synthesised at levels 10 times higher then for the mature enzyme alone [108]. It seems likely that ubiquitin does have a translatory activating function within UbCEPs, however, the question of whether it also acts as a molecular chaperon thereafter is yet to be resolved.

Section 1.2 Ubiquitin Structure

The primary structure of ubiquitin is uniquely conserved across eukaryotes; its sequence is identical among animals, and differs from this in only a few sites in yeast [10] and plants [12]. Ubiquitin contains one histidine and one tyrosine, but no tryptophan or cysteine residues, hence there are no structure stabilizing disulphide bridges. However, the observation that human and yeast ubiquitin's are indistinguishable in their ability to sustain ubiquitin dependant protein degradation in mammalian reticulocytes [109] strongly suggests the same well defined tertiary structure is common to both.

1.2.1 The Secondary and Tertiary Structure

In an early investigation, Lenkinski and co-workers noted that the 1D NMR spectrum of ubiquitin changed very little over the temperature range 23 - 80 ^{o}C and the pH range 1 - 8 [110]. Although carried out on a relatively, at least by today's standards, insensitive 270 MHz spectrometer, these results suggest ubiquitin possesses a very stable tertiary structure, a conclusion supported by its resistance to digestion by many enzymes [111].

General information on the secondary structure of ubiquitin was obtained by Goldstein and co-workers, who performed a series of physico-chemical experiments on the protein [112]. Their interpretation of the far ultra-violet circular dichroism (CD) spectrum of ubiquitin, suggested the protein contained little organized secondary structure. This conclusion was based on similarities between a theoretical spectrum (calculated for 6% α-helix, 10% β-sheet and 84% random coil) with that of ubiquitin. However, agreement between theoretical and experimental curves rapidly breaks down below 210 nm, raising doubts about the validity of such a direct comparison. No significant changes in the CD spectra were observed between pH 3 and 12, which in keeping with the previously described NMR study implies ubiquitin has a stable conformation.

In the same study, both ultra-violet absorption spectroscopy and fluorescence spectroscopy were used to probe the environment of the single tyrosine residue in ubiquitin (consequently, it is far easier to interpret fluorescence data from proteins containing single fluorophores than for those with multiple fluorophores [113]). Spectrophotometric titration of the tyrosine revealed it has a pKa of 11.1, a high value usually indicating the phenolic group of tyrosine is involved in hydrogen bonding. Furthermore, the fluorescence of the tyrosine is greatly quenched relative to a free tyrosine amino acid standard. This type of quenching is commonly caused by other amino acids in the protein being in the immediate vicinity of the fluorophore [113]. Both solvent perturbation absorption and iodide quenching of the fluorescence indicated that the tyrosine is partially buried within the ubiquitin structure. Finally, other fluorescence studies suggested that the protein is fully denatured in 7M guanidine.

The crystallization and preliminary X-ray investigation of bovine ubiquitin was first reported by Cook and co-workers in 1979 [114]. The same group later published the full crystal structure of human erythocytic ubiquitin at 2.8 Ang. resolution [3], and ultimately refined this to 1.8 Ang. resolution [115]. These X-ray studies have revealed ubiquitin to be an extremely compact globular protein, in which around 87% of the peptide backbone is involved in hydrogen bonded secondary structure. Oat and yeast ubiquitin sequences have also been crystallized, and analysis of these reveals they have similar tertiary structures to that of the animal form [116]. A ribbon representation of the tertiary structure of ubiquitin is shown in figure 1.9.

Examination of its crystal structure reveals ubiquitin contains all the secondary structure units commonly found within proteins, the small size of ubiquitin results in these being tightly packed together forming a very compact tertiary structure. The most salient of these are 3.5 turns of α-helix involving residues 23 - 34, a short length of 3_{10} helix incorporating amino acids 56 - 59 and a five stranded mixed β-sheet. In addition, there are seven reverse turns (types I-III) linking the various helices and components of the sheet together. The left handed twist of the β-sheet forms a cavity into which fits the α-helix, thus forming the basis of the pronounced hydrophobic core exhibited by the protein. It is a combination of this hydrophobic core and hydrogen bonding between helices and turns, that is thought to account for the remarkable stability of the molecule to ranges of temperature and pH [110].

One of the most prominent features of the tertiary structure is the C-terminal region. Amino acids 73-76 clearly protrude from the core of the molecule, indeed, the C-terminus is often referred to as the "stinger" in ubiquitin's tail. The absence of any surrounding structure would suggest the C-terminal region has considerable freedom of movement, this may have some significance given that much of the protein's biological function involves this region.

One curious feature in the structure is the unusual hydrogen bond which occurs between the sulphur atom of Met-1 and the peptide NH group of Lys-63. This effectively buries the amino terminus, thus rendering it inaccessible, a property which may explain the protein's apparent immunity from degradation via its own pathway. Moreover, given that ubiquitin can be synthesised as a polymeric precursor in which individual units are linked in a head to tail manner [8-12], the N-terminal region must undergo considerable conformational changes after cleavage of the precursor into mature ubiquitin molecules.

Another significant hydrogen bond occurs between the phenolic group of Tyr-59 and the backbone nitrogen atom of Glu-51. This interaction, which explains the tyrosine's atypical physico-chemical properties [112], helps to stabilize the loop occurring between a reverse turn (residues 51-54) and the 3_{10} - helix (residues 56-59). The tyrosine appears to reside at a "hinge-like" position in the protein, that is to say, the loop in which it is located practically ensures that the final component of

the β-sheet (residues 64-72) is oriented in the correct way. Destabilization of this loop, essentially loss of the Tyr-59 - Glu-51 hydrogen bond, may result in the final strand of the β-sheet not being formed, indeed, one might speculate that another conformation would ultimately be adopted by the protein (see section 1.2.2).

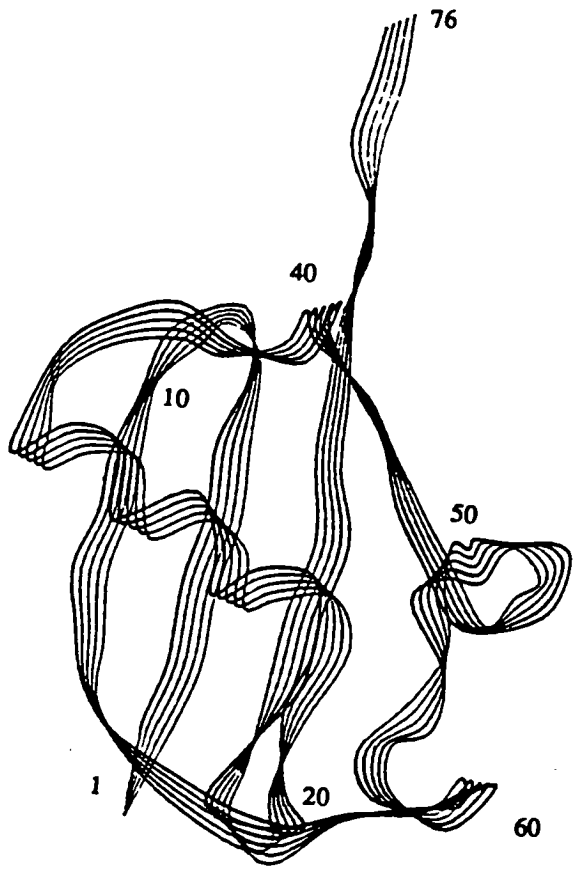


Figure 1.9 Ribbon Structure of Ubiquitin

One final piece of information that can be abstracted from the crystal structure, is the availability of lysine residues for isopeptide bond formation. Multi-ubiquitination of substrate molecules involves the formation of an isopeptide bond between the ε-amino group of lysine in one ubiquitin and the C-terminal of the next one in the chain. Of the seven lysines in ubiquitin, two (Lys-11 & Lys-27) are completely inaccessible. Three of the remaining five (Lys-6, Lys-33 & Lys-63) are fully exposed on the surface of the protein and form no intra-molecular contacts.

Lysine 29 and lysine 48 are both on the surface of the protein, but do have hydrogen bonds between their ε-amino groups and peptide carbonyl groups. Therefore, it appears that lysines 6, 33 & 63 should be predisposed to isopeptide bond formation with residues 29 & 48 also being candidates, but to a lesser extent.

Until recently, X-ray crystallography offered the only approach to tertiary structure elucidation of small proteins, however rapid advances in both methodology and spectrometer technology, have resulted in nuclear magnetic resonance spectroscopy becoming a fully complementary technique. The principle advantage NMR has over X-ray crystallography is that it provides the conformation of the protein in solution, its natural environment, and this allows more dynamic investigations to be carried out, ie protein folding.

In the past the sheer complexity of a protein's 1D NMR spectrum made comprehensive proton resonance assignment impossible. The advent of powerful spectrometer magnets, along with the two-dimensional NMR experiment (2D NMR), has not only made complete proton assignment a reality, but has also allowed secondary and tertiary structures to be determined. Since its conception in the early seventies by Jeener and Ernst [117], 2D NMR has progressed with phenomenal speed, indeed today there exist a number of such experiments, each delivering different, yet complimentary, information on the connectivities of individual proton resonances. Application of these techniques to the sequential assignment of proton resonances in small proteins, was introduced and developed by Wüthrich and co-workers [118-120]. A comprehensive treatment of 2D NMR studies on biomolecules, which is outside the scope of this text, is given in Wüthrich's seminal book on the subject [121].

Two papers, both describing the complete ^{1}H NMR assignment and subsequent secondary structure analysis of ubiquitin, were published simultaneously in 1987 [4,5]. In each case, a combination of **D**ouble **Q**uantum **F**ilter **C**orrelated Spectroscop**Y** (DQF-COSY [122]), **D**ouble **Q**uantum **F**ilter **R**elayed Spectroscopy (DQF-RELAY [123]), **T**otal **CO**rrelated Spectroscop**Y** (TOCSY [124]) and **N**uclear **O**verhauser and **E**xchange Spectroscop**Y** (NOESY [125]) were used to assign the observable proton resonances. Secondary structure was identified using characteristic through space interactions observable in the NOESY spectrum. For example, in helices there are typically strong crosspeaks between adjacent amide

protons (denoted dNN), also because of the helical shape, medium range interactions (eg, $d\alpha\beta_{[i,i+3]}$) would be expected[121]. The secondary structure of ubiquitin in solution (derived from all of these sequential, medium and long range NOESY crosspeaks) is in remarkable agreement with its crystalline counterpart. In their assignment, Di Stefano and Wand identified the long α-helix and short 3_{10} - helix, the five stranded mixed β-sheet and five of the seven β-turns [5]. Similarly, Weber and co-workers' [4] secondary structure assignment was completely consistent with the crystal structure [115]. Although no attempts were made to elucidate the tertiary structure, clearly both these studies strongly point to ubiquitin having an identical overall conformation in solution and crystal phases.

The tertiary structure of a protein in solution can also be derived from 2D NMR data [121]. Distance-Geometry (DG) constraints, extracted from the appropriate 2D NMR experiment, are used to optimize the structure by way of molecular mechanics or dynamics calculations. Distance constraints (by constraints, approximate values are meant) can be obtained from NOESY crosspeak intensities, whereas geometrical information (ie torsion angles) can be acquired from J-coupling values between nuclei by way of the Karplus relationship [126]. Weber and co-workers used 700 constraints to generate a distance-geometry derived tertiary structure for ubiquitin [127]. Two different DG programs were used, AMBER [128] and DSPACE [129], both of which gave essentially the same result; ubiquitin has an energy minimized solution conformation very close to its X-ray structure (there are minor changes in side-chain positions). It is noteworthy that neither of these DG programs use the X-ray structure as a starting point for minimisation.

1.2.2 Structure Function Studies on Ubiquitin

To date, only a few studies have addressed the question of structure-function correlations in ubiquitin. The first, and by far the most comprehensive of these, was published in 1987 by Ecker and co-workers [130]. In this study, the ubiquitin conformation was conveniently divided into three general categories.

1. The hydrophobic core.

2. The protein surface.

3. The C-terminal region.

Examination of the ubiquitin tertiary structure using molecular graphics, allowed important interactions to be identified within each category, and from these appropriate mutations were conceived (see table 1.2).

PROTEIN STRUCTURE	LOCATION OF CHANGE	*in vitro* BSA CONJUGATION	EFFECT ON STRUCTURE
Animal ubiquitin		100 %	
Pro19 → Ser, Ala28 → Ser, Glu24 → Asp	Surface	100 %	Minimal
Pro19→Ser	Surface	100 %	Minimal
Leu67→ Asn, Leu69 →Asn	Core	0 %	Severe
Gly76 →Ala	Tail	0 - 10 %	Minimal
Leu73→deleted	Tail	0 %	Minimal
Leu 73→deleted, Arg72→Ser	Tail	0 %	Minimal
Tyr59→Phe	Surface	70 - 100 %	Slight
His68→Lys	Surface	30 %	Slight
Tyr59→Phe, His68→Lys	Surface	30 %	Slight

Table 1.2 Ubiquitin Mutants and their Biological Activities

These mutants, some containing multiple changes, were then synthesised by construction of a plasmid encoding a synthetic ubiquitin gene into which eight restriction enzyme sites had been engineered, thereby facilitating the range of mutations required. After site directed mutagenesis had been performed, the plasmid was expressed in E. coli and the resulting mutant ubiquitin purified by ion exchange chromatography.

The effect these mutations had on the protein's structure, was determined by NMR spectroscopy. In addition, each mutant ubiquitin was assayed for its ability to be recognized and activated by a ubiquitin activating enzyme (E1), as well as its ability to support protein conjugation in an *in vitro* reticulocyte fraction II degradation assay.

Two ubiquitin mutants with conservative surface substitutions, that is to say the triple mutant Pro19 - Ser, Ala28 - Ser & Glu24 - Asp and the single site mutant Pro19 - Ser, had both E1 recognition and protein conjugation activities identical to the wild type protein. Furthermore, other than at the point of mutation, no changes were observed in the 1D NMR or in the DQF-COSY 2D NMR in either of these mutants relative to wild type ubiquitin. This indicates that these mutations have very little effect on the protein's overall structure. Therefore, it seems ubiquitin is quite tolerant to these rather guarded mutations on its surface.

Two surface mutations with some damaging effect upon both E1 recognition and *in vitro* protein conjugation activity, were the Tyr59 - Phe and the His68 - Lys substitutions. In the light of this, it is somewhat surprising that neither of these mutants had any significant changes in their 2D NMR spectra, again suggesting very little structural reorganization had occurred relative to the native form. One apparently contradictory observation was that the Tyr59 to Phe mutant was extremely insoluble. This is often associated with exposure of hydrophobic residues, which in this case could be caused by disruption of the tyrosine containing loop.

The importance the hydrophobic core has upon the ubiquitin structure, was investigated using the double mutant Leu67 - Asn, Leu69 - Asn. These two leucine residues, both components of the hydrophobic core, can be mutated to asparagine without creating any steric problems. Of course, such mutations would lead to the core being considerably less hydrophobic. In fact, this double mutant had absolutely no E1 recognition activity or protein conjugation activity, and perhaps not surprisingly, the NMR data indicated the native structure of ubiquitin had been devastated by these two changes. The introduction of two weakly hydrophilic side-chains into the hydrophobic core, would create energetically unfavourable interactions which may well be enough to prevent the protein from folding. Clearly the hydrophobic core of ubiquitin is very important to its structure and, consequently, the protein is very intolerant to drastic changes in this region.

The final aspect of the ubiquitin structure to be investigated was the C-terminal region. Three "tail mutants" were constructed (Gly76 -Ala, Leu73 - Δ & Leu73 - Δ, Arg72 - Ser) all of which were completely inactive in both the E1 recognition assay and the protein conjugation assay. As one would expect, NMR indicated that

mutations in the tail had no effect on the bulk structure of protein. The first step in the ubiquitin pathway is activation of the C-terminus by the E1 enzyme. Mutations in the tail lead to loss of activity, implying the E1 enzyme recognizes specific features within this region, as well as the protein conformation as a whole.

Although ubiquitin alone is regarded as an extremely stable protein, over the years there has been some speculation over whether it undergoes conformational changes upon conjugation to other proteins. Wilkinson and Mayer found that lowering the dielectric constant of a ubiquitin solution, by addition of various alcohols, was accompanied by a dramatic increase in the helical content of the protein [131]. For example, lowering the dielectric constant from 80 to 45, resulted in ubiquitin having a circular dichroism spectrum consistent with 50% helical structure. Indeed, it is commonly accepted that addition of alcohols can increase the helical content of peptides and proteins [132,133]. In a related study, Williams and co-workers claim that at low pH, addition of methanol to a ubiquitin solution brings about a stable, partially structured form of the protein [134]. Ubiquitin is completely transformed into this form in 60% methanol/ 40% water at pH 2.0, and detailed characterization of this structure was carried out by 2D NMR. The α-helix and first two strands of the β-sheet remain remain intact in this novel form, however, the turn rich area of the protein, which includes the β-turn - 3_{10}-helix loop, appears to be destroyed. Unlike the preceding CD study, this NMR analysis suggests the helical content actually decreases in the more hydrophobic environment, an unexpected observation which has yet to be resolved. Clearly though, exposure to a more hydrophobic environment does induce changes in the ubiquitin structure. *In vivo*, ubiquitin could encounter a localised hydrophobic environment in the vicinity of a cellular membrane, this might be enough to trigger the above conformational change. Also, conjugation to particularly hydrophobic proteins might induce a switch to this other form.

In an attempt to elucidate the importance of conformational mobility in ubiquitin mediated proteolysis Ecker and co-workers constructed, expressed in E. coli and purified a number of cysteine containing mutants of ubiquitin [135] (see table 1.3).

MUTANT STRUCTURE	No of Cys	Activity
Bovine Ubiquitin	0	100 %
Phe4 - Cys	1	90 %
Thr14 - Cys	1	68 %
Thr66 - Cys	1	110 %
Phe4 - Cys, Thr14 - Cys	2	100 %
Phe4 - Cys, Thr66 - Cys	2	30 %
Phe4 - Cys, Thr14 - Cys, Thr66 - Cys	3	20 %

Table 1.3 **Cysteine Containing Mutants of Ubiquitin *In vitro* Protein Conjugation Activities are shown relative to Bovine Ub**

Three of these mutants contained single X-Cys substitutions, while the remainder contained two or more mutations, the latter facilitating disulphide bridge formation. All six mutants were recognized and hence activated by the E1 enzyme, however, there were striking differences in their *in vitro* protein conjugation activities. The three mutants containing single cysteine residues, as well as the mutant containing a disulphide linking the first two strands of the β-sheet (Phe4 - Cys, Thr14 - Cys), were found to be as active as wild type ubiquitin in the conjugation assay. On the other hand, the two mutants containing the very constricting disulphide linking the amino and carboxyl termini (Phe4 - Cys, Thr66 - Cys and Phe4 - Cys, Thr14 -Cys & Thr66 - Cys) were only 20 -30 % as active as wild type ubiquitin.

The tertiary structure of the 4/66 disulphide was determined using 2D NMR followed by distance-geometry calculations (AMBER & DSPACE). In solution, the 4/66 mutant was found to have a left-handed disulphide geometry, but more importantly the overall conformation of the molecule was more or less identical to wild type ubiquitin. Together these results suggest that initial activation of ubiquitin by E1 does not involve any conformational change (all the mutants were activated), however, full protein conjugation activity requires a certain amount of conformational mobility. If ubiquitin is conformationally constrained, as in the 4/66 disulphide, then protein conjugation and hence degradation decreases perhaps

because ubiquitin can no longer achieve a conformation that is active in signaling proteolysis.

The importance of the unusual hydrogen bond between the sulphur atom in methionine and the peptide NH of Lys-63, has been investigated by Breslow and co-workers [136]. Chemical oxidation of Met-1 to the sulphoxide or to the sulphone, was found to destabilize the ubiquitin structure resulting in the protein becoming more susceptible to conformational change when, for example, the di-electric constant of the solvent is lowered. Interestingly, oxidation of the sulphur in methionine results in ubiquitin being slightly more active in protein degradation. This reinforces the contention that ubiquitin requires a certain amount of conformational freedom to fulfil this function. Cyanogen bromide cleavage of the final methionine affords a truncated ubiquitin species which completely unfolds below pH 4.0, it is also much less active in its protein degradation function. The authors suggest that the oxidised Met species may still have a hydrogen bond (although weaker) between the suphoxide or sulphone oxygen atom and NH of Lys-63. No such hydrogen bond can form in the des-Met species and consequently removal of the methionine is a far more damaging modification.

1.2.3 The Ubiquitin Folding Pathway

There are two fundamental questions that can be asked about the tertiary structure of a protein:

(i) Why did it adopt this particular conformation in the first place?

(ii) Given that a particular fold is somehow inherent in a protein's sequence, just what is the mechanism involved in going from the unfolded post-ribosomal state to this final conformation, if indeed this is a 2-stage process?

1.2.3(i) Protein Structure in General

In addressing the first question surely what we have to ask is, what factor or factors stabilize a given tertiary structure over every other? The identity of these factors has never really been in dispute, it has long been agreed that the two principle forces governing protein stability are the hydrophobic effect and hydrogen bonding. A protein's native conformation should be that which contains the greatest possible stabilizing contributions from these two elements. However, destabilizing

contributions from the loss of internal rotations and the increase in torsional strain associated with folding, result in the overall net stability being rather small (typically 5 - 20 kcal mol^{-1}).

Over the years there has been some controversy over which of these two stabilizing determinants has the major input. As long ago as 1936, Pauling and Mirsky contended that hydrogen bonding was the more significant of the two [137]. This contention remained more or less unchallenged until the 1960's, when the work of Kauzmann [138], and later Tanford [139], suggested that in model compounds at least, the hydrophobic effect was the most important in determining the three dimensional structure of a protein.

In recent years, site directed mutagenesis has provided researchers with the experimental means to investigate the relative importance of the hydrophobic effect *vs.* hydrogen bonding in protein tertiary structure. While opinion is undeniably split over their relative contributions, the majority of researchers believe the hydrophobic effect to be the dominant, and have accordingly designed their experiments to show this. These experiments generally involve the mutation of hydrophobic residues within a protein's core to smaller hydrophobic residues such as alanine. The stability of the resulting mutants are then measured relative to the wild type protein. Unfortunately, such experiments have not afforded any unifying data, for example Matthews and co-workers have found that Leu to Ala mutations in the core of lysozyme can lower the free energy of stabilization (ΔG_{H2O}) by between 2.0 and 8.0 kcal mol^{-1}, depending upon the position of the substitution [140]. Fersht and co-workers [141,142] have similarly found that the effect a mutation has on the stability of a protein, in this case barnase, is very much dependant upon its position within the core (their results suggest a loss in ΔG_{H2O} of between 1.0 and 1.6 kcal mol^{-1} per methylene group removed from the core).

One of the leading exponents of the hydrogen bonding school of thought, is Williams. Based upon his work on small molecules he suggests that hydrogen bonding may well make a much underestimated contribution to protein stability [143,144]. Model systems have revealed that a single hydrogen bond can provide a free energy of stabilization of between 3 and 5 kcal mol^{-1}, most of which comes from an entropic factor associated with the release of bound water molecules. In a recent paper, Pace and co-workers have used site directed mutagensis to probe the

importance hydrogen bonding has to the conformational stability of ribonuclease T1 [145]. Deletion of a series of intra-molecular hydrogen bonds, by construction of the necessary mutants, was found to destabilize the overall structure by on average 1.3 kcal mol^{-1} per bond deleted. Given the number of such bonds within the protein, this group conclude that hydrogen bonding makes a comparable contribution to the hydrophobic effect in stabilizing its structure.

The apparent discordance of opinion and results within this whole area makes it difficult to abstract any unifying principles. This lack of consistency is understandable since every protein has its own unique contributions from both hydrogen bonding and the hydrophobic effect and, therefore, any attempt to construct a detailed set of general rules will be extremely difficult.

1.2.3(ii) Protein Folding in General

It has long been accepted that there are just too many conformations of an unfolded protein for its native conformation to be adopted by random processes, statistically such a mechanism would require a ridiculously long timescale. If we, therefore, rule out random processes, this leads us to conclude that protein folding is an inevitable consequence of the primary sequence, and that all the necessary information required for this phenomenon is somehow contained therein.

The work of Lim and Sauer on the λ receptor protein has led them to conclude that the particular connectivities between hydrophilic and hydrophobic residues in the sequence, limits the number of ways the protein can arrange itself so as to position the majority of the hydrophobic residues on the inside and hydrophilic residues on the outside [146]. In evolutionary terms, it has been suggested that all present day protein structures descend from a finite number of simple degenerate patterns which nature long ago discovered were sufficient to give a basic globular protein scaffold [147].

It is now generally accepted that there exist protein folding pathways with well defined intermediates, so called molten globular states [148]. While precise structural analysis of these intermediates would undoubtedly represent the first step to understanding protein folding, it is only relatively recently that techniques to do so have become available. These novel NMR techniques all rely on the fact that amide protons involved in hydrogen bonded secondary or tertiary structure, exchange at a

slower rate than those not involved in such bonding. If a protein is unfolded in a D_2O denaturant solution, then all backbone amide protons will be replaced by deuterons. Rapid dilution of the protein solution in D_2O allows folding to proceed. However, if after certain times intervals the refolding protein is briefly exposed to H_2O at basic pH, then only those deuterons not involved in hydrogen bonding will exchange with protons. Consequently, only those protons not involved in hydrogen bonding in the intermediate will be observed in the ^1H NMR spectrum of the refolded protein.

This type of stop-flow NMR experiment was first applied to bovine pancreatic trypsin inhibitor (BPTI), a relatively small 6 kDa protein, although only a 1D NMR analysis was carried out [149]. More recent folding studies, all involving powerful 2D NMR approaches, have been carried out on cytochrome C [150], barnase [151], lysozyme [152] and ribonuclease A [153]. These studies have all shown that secondary structures, such as helices, are formed very early in the folding pathway and that the generation of tertiary structure hydrogen bonds occurs at a later stage. This is broadly in keeping with the kinetic circular dichroism results of Kuwajima, which also indicate secondary structure is formed very early in the folding pathways of several proteins [154].

Clearly this H-D exchange technique does have its limitations, for example, it is completely insensitive to hydrophobic interactions and can only be applied to proteins small enough to have fully assigned ^1H NMR spectra. These limitations aside, the technique has furnished us with the only tangible evidence for the existence of folding intermediates and its continued application to other proteins is currently the only way of assessing the generality of such molten globule states.

1.2.3(iii) Stability of the Ubiquitin Structure

Ubiquitin, as we have seen, has a very condensed conformation containing a high degree of hydrogen bonded secondary structure and a pronounced hydrophobic core. In asking which of these two elements contributes most to the stability of the ubiquitin structure, it is necessary to examine the effects specific mutations make on this conformation. Although not ideal for this purpose, the work of Ecker and co-workers (discussed in section 1.2.2) can certainly give us an insight into the relative importance of these two fundamental components in stabilizing the ubiquitin structure [130]. Very briefly, this group found the ubiquitin conformation

to be more tolerant to mutations disrupting hydrogen bonding than to mutations disturbing the hydrophobic core of the protein, perhaps implying the latter conveys the major stabilizing contribution. Neither this nor any other group has tried to quantify the contribution made by the hydrophobic effect to the overall free energy of stabilization of ubiquitin.

1.2.3(iv) The Ubiquitin Folding Pathway

The folding pathway of ubiquitin has recently been investigated using the pulse-labelling 2D NMR approach described in section 1.2.3(ii). In this study, Briggs and Roder found that amide protons in the α-helix and β-sheet, as well as those involved in hydrogen bonds between these two regions of secondary structure, become protected from H-D exchange very quickly (8 ms) after refolding is allowed to begin [155]. This indicates that these two elements of secondary structure rapidly form and associate with each other during folding, in fact the authors suggest that the formation and association of the α-helix and the β-sheet is a co-operative folding event. A much later folding event in the pathway seems to be the formation of the surface loop containing the 3_{10}-helix and the reverse turn incorporating residues 59,61 &69. These residues have much slower rates of protection from H-D exchange, indicating their loop stabilizing hydrogen bonds are correspondingly as slow in forming.

Interestingly, this folding intermediate appears to remarkably similar to the alcohol induced partially structured state of ubiquitin described by Williams and co-workers [134]. The similarities are quite remarkable, both contain the α-helix and β-sheet and both apparently lack the afore mentioned surface loop. Perhaps the most intriguing aspect of the partially structured state proposed by Williams is that it is effectively produced going in the opposite direction (folded to unfolded) to the refolding experiment of Briggs and Roder, implying the same intermediate is associated with both the denaturing and renaturing of ubiquitin [155].

1.3 Peptide Synthesis

1.3.1 General

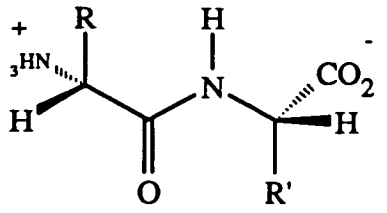
The term "peptide" is derived from the greek word "pepsis" which means digestion (of proteins). Consequently, peptides are typically low molecular weight compounds, albeit low refers to their weight relative to proteins. At exactly what size a peptide ceases to be a peptide and becomes a protein is really a matter of taste, but 100 amino acids is generally recognized as being the cut-off point. The relationship between peptides and proteins goes far beyond this simple arithmetic connection, in that the majority of peptides actually originate from proteins which have been processed by specific enzymes. As with their protein counterparts, peptides are employed by nature to execute all manner of biological functions.

Cellular assembly of peptides and proteins involves two distinct steps. DNA, the genetic blueprint for all peptides and proteins, is initially transcribed into mRNA which then acts as the template for their subsequent ribosome-mediated construction. Within the cell these processes are performed with great speed and fidelity, thus, accounting for the proliferation of cellular proteins. The peptide chemist, attempting to artificially carry out something that nature performs with almost unerring efficiency, is confronted by a whole host of problems that make chemical synthesis very laborious in comparison. Given the apparent difficulties of peptide synthesis, it is perhaps surprising to learn that Emil Fisher reported the synthesis of the dipeptide, glycylglycine, as long ago as 1901 [156]. The following sixty years saw a gradual understanding of the fundamental requirements and problems specific to peptide construction, and an appreciation of these allowed Merrifield to conceive solid phase peptide synthesis (SPPS) in 1963 [157].

The field of peptide chemistry is now such a large one that it would be impossible to cover every facet of the subject within the limited space available, several reviews and books cover the subject thoroughly (for example refs. 158, 159). However, there are selected crucial topics which should be discussed.

1.3.2 Peptide Synthesis - A Chemical Challenge

The majority of the chemical obstacles confronting the chemist can be adequately illustrated by considering the condensation of two α-amino acids to give the dipeptide (**1**).



(1)

Figure 1.10 A Dipeptide Illustrating the Trans-Peptide Bond

As they stand, the two amino acids will simply not react with one another to form a peptide bond. This lack of nucleophilicity and electrophilicity can be attributed to the zwitterion nature of amino acids. To overcome this hurdle, a dehydrating agent, the so called "coupling agent" must be added to the mixture, this allows formation of the peptide bond to take place. If the two amino acids are different in any way, then clearly more than one product will be obtained (see figure 1.11).

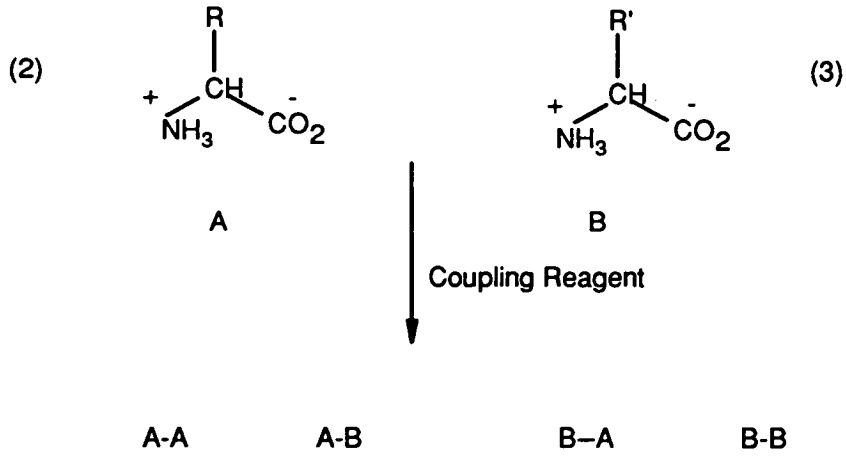


Figure 1.11 Ambiguity of a Simple Coupling Reaction

Moreover, this problem is compounded by the observation that many of the side-chains of amino acids contain functional groups (eg 4 & 5 figure 1.12) that can also react, increasing the number of possible products still further.

Aspartic Acid (4) Lysine (5)

Figure 1.12 Reactive Side-Chains of Two Amino Acids

The codon-anticodon system used by nature prevents such random processes from occurring, however, the chemist has to protect certain parts of the molecule from reaction in order to achieve complete specificity. Figure 1.13 illustrates a suitable protection strategy for the reaction of (**4**) with (**5**).

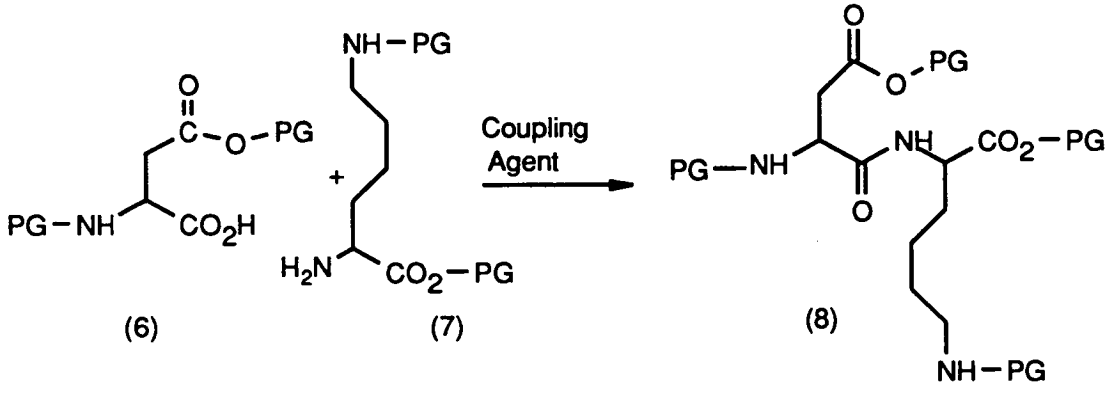


Figure 1.13 Unambiguous Condensation of Aspartic Acid and Lysine

Until now only the construction of a dipeptide has been considered. If, however, we wish to extend the peptide any further, whilst retaining absolute specificity in each coupling step, then a level of orthogonality must be incorporated into the protection strategy. Clearly, either the N-terminal or the C-terminal protecting group will have to be removed, whilst leaving the remainder unaffected.

1.3.3 Activation and Coupling Methods

Activation of the α-amino group can be brought about by, for example, conversion to the hydrazide [160]. This then facilitates the coupling to the carboxylic acid group of the next amino acid, as shown in figure 1.14.

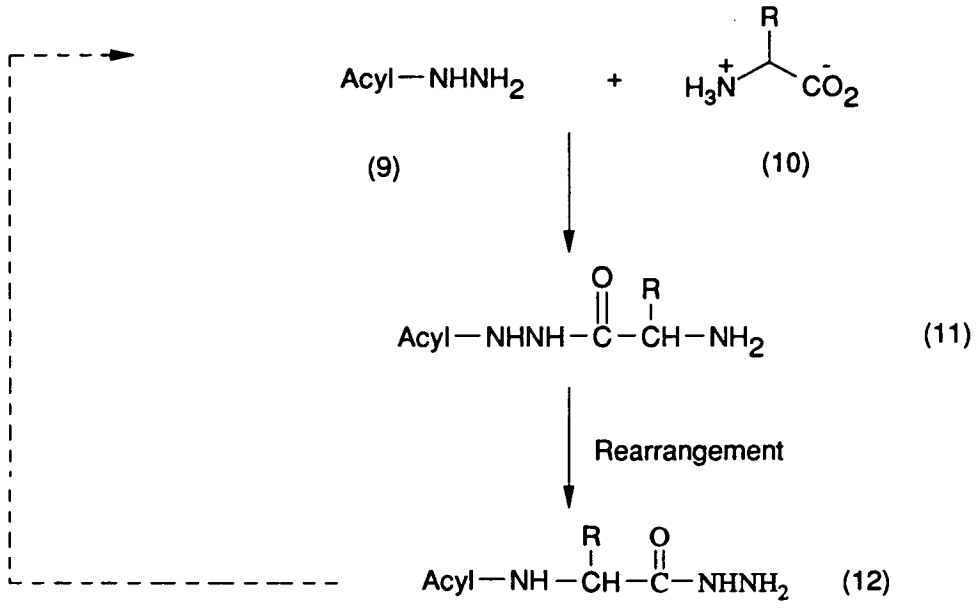
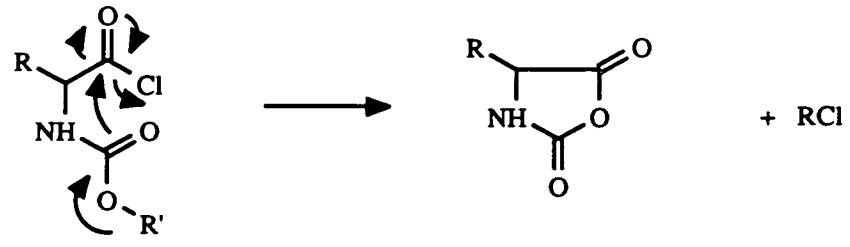


Figure 1.14 Hydrazide Activation in Peptide Synthesis

Although this procedure involves some rather elegant chemistry, it does require somewhat elaborate side-chain protection, and for this reason is not commonly used.

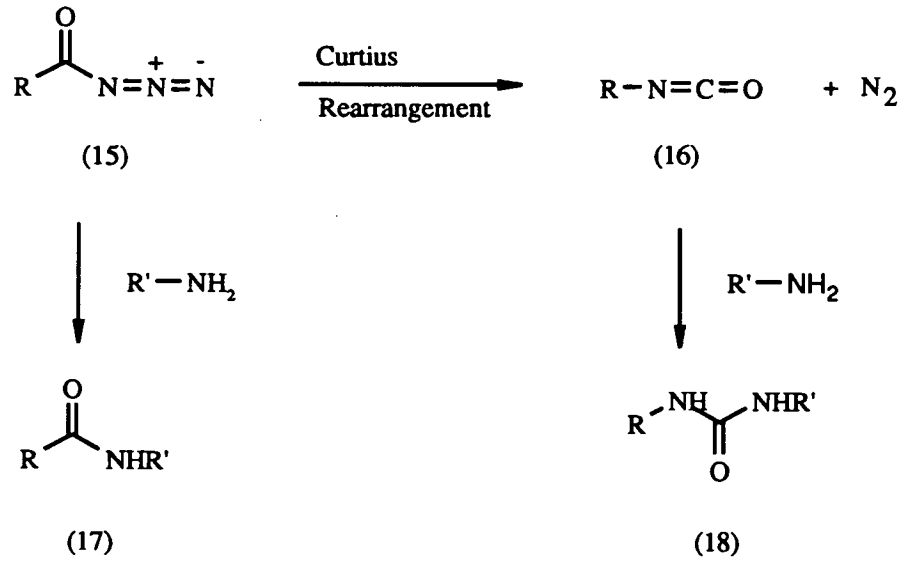
A far more convenient approach is to activate the carboxyl group to nucleophilic attack. Replacement of its hydroxyl function with any electron withdrawing group (X), will substantially increase the electrophilicity of the carboxyl group. Generation of the acid chloride is an obvious route to carboxyl activation and indeed this method was used in Fisher's early attempts at peptide synthesis [161]. Acid chlorides, though powerful acylating agents, do have one serious limitation. Under the reaction conditions, the commonly used urethane N^{α} protecting group becomes involved in a process which gives rise to N-carboxyanhydrides (**14**), these highly reactive amino acid derivatives are generally unsuitable for peptide synthesis due to instability and loss of chiral integrity.



(13) (14)

Figure 1.15 Conversion of a Protected Amino Acid Chloride into an N-Carboxyanhydride

The use of acid azides (**15**) in carboxyl activation also dates back to the turn of the century [162]. While this mode of activation, pioneered by Curtius, is generally of greater practical importance than the acid chloride approach, it is limited by the tendency of the acid azide to rearrange to the isocyanate (**16**).

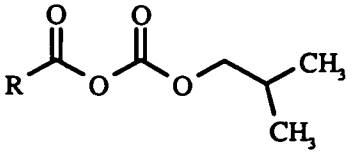


(15) (16) (17) (18)

Figure 1.16 The Acid Azide Method

Given that amines are readily acetylated using acetic anhydride, it is perhaps surprising that amino acid anhydrides were not used as coupling reagents from the outset. Several mixed anhydrides have enjoyed various levels of success as activating agents (reviewed in ref. 159). All of these can be conveniently prepared by treatment of the protected amino acid with condensing agents such as

diisopropyl- [163] or dicyclohexylcarbodiimides [164], usually referred to as DIC and DCC respectively. By far the most widely used of these are the isobutylcarbonic acid mixed anhydrides (**19**). These have the advantage of producing easily removed carbon dioxide and isobutanol as by-products of the acylation reaction [165].



(19)

Figure 1.17 Isobutylcarbonic Acid Mixed Anhydride

Mixed anhydrides can of course react with amines to give two acylated products. This regioselectivity problem can be reduced by increasing the steric bulk of the non-amino acid component of the anhydride. Furthermore, the use of mixed anhydrides in which the amino acid component is more electrophilic than the other constituent (eg carboxylic-phosphinic mixed anhydrides [165]) results in only the desired acylation being obtained. However, phosphinic anhydrides are known to be rather unstable, thereby limiting their application. Symmetrical amino acid anhydrides (**21**) possess no such limitation and have, thus, found widespread use. As with mixed anhydrides, they can be easily prepared from protected amino acids using DIC or DCC (**20**).

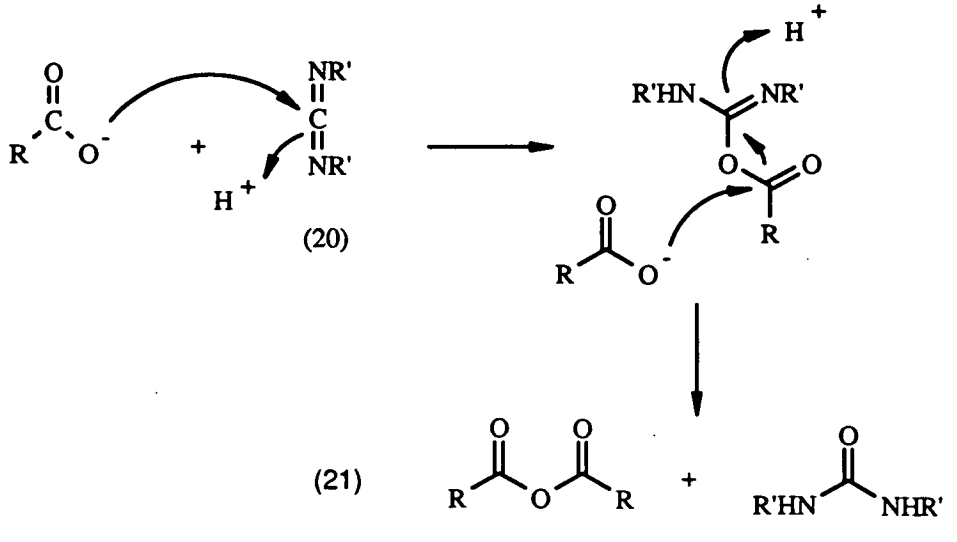
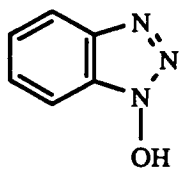


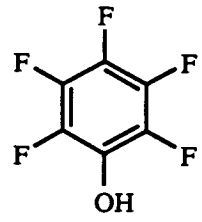
Figure 1.18 Symmetrical Anhydride Formation

Possibly the only disadvantage of the symmetrical anhydride approach, is that for every amino acid molecule coupled to the growing peptide one is lost as a by-product of the reaction, thus making the procedure expensive.

The active ester approach to coupling creates no such waste of expensive amino acids. Esters of amino acids only become sufficiently active when the alcohol component possesses a negative inductive effect, thereby rendering the carbonyl carbon more electrophilic. Ortho- and para-nitrophenyl esters were among the first widely used active esters [166]. Other, more recently introduced, active esters can be obtained from DIC/DCC mediated condensation of 1-hydroxybenzotriazole (**22**)[167] or pentafluorophenol (**23**) [168] with suitably protected amino acids.



(22)



(23)

Figure 1.19 Two Commonly Used Alcohols in Esterification

1.3.4 N^{α} Protecting Groups

The prerequisites of an effective N^{α} protecting group would be as follows:

1. It must mask the amino group of an amino acid or peptide during the acylation step. This will ensure the correct peptide bond is formed.

2. If the peptide is being synthesised from the C to N terminal, then it must also be removable under relatively mild conditions which leave all other protecting groups intact.

Acylation of the α-amino group is an obvious, but ultimately unsuitable means of protection. N^{α}-acyl protecting groups are not only extremely difficult to remove, but also lead to oxazolone (**24**) formation during the activation step. Base abstraction of the α-CH proton after oxazolone formation leads to racemisation of the amino acid, as illustrated in figure 1.20.

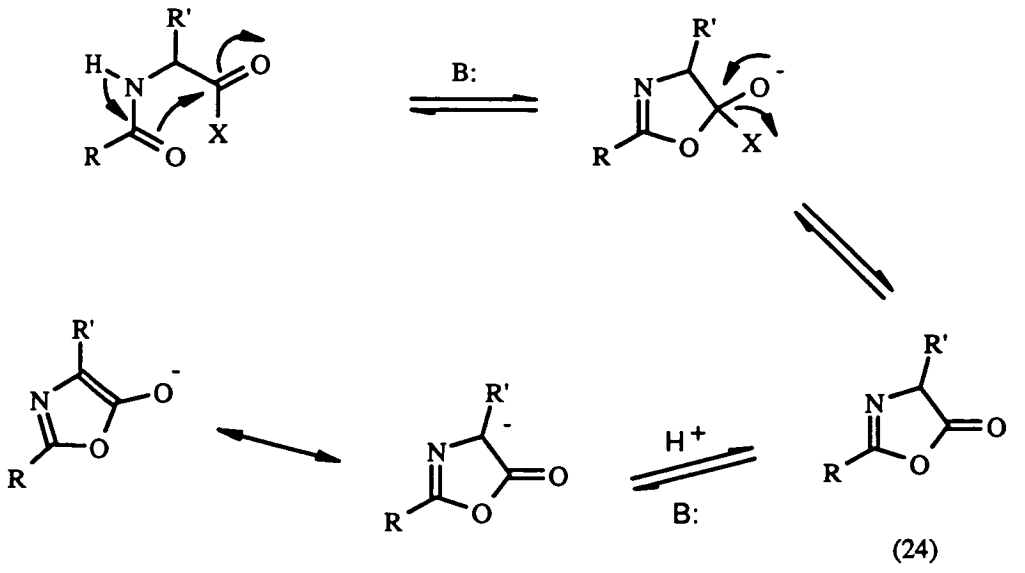


Figure 1.20 Racemisation of N^{α}-Acyl Protected Amino Acids

The observation that carbamic acids (**26**) spontaneously decarboxylate to give a free amino group, serves as the basis on which all modern-day N^{α} protecting groups are designed. Notional esterification of the carbamic acid gives rise to a urethane (**25**) which offers stable protection until its ester component is cleaved (see figure 1.21).

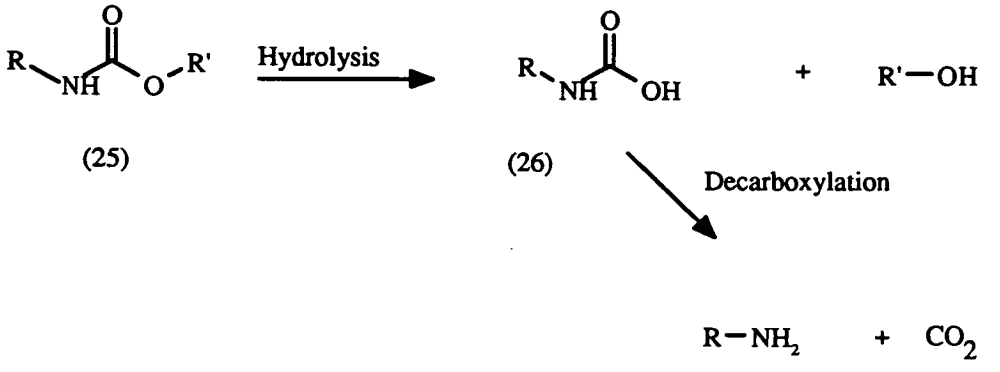


Figure 1.21 Urethane Deprotection

The introduction in 1932 of the benzyloxycarbonyl or Z group (**27**) [169] not only signified the first successful application of a urethane to N^{α} protection, but also stands as a milestone in peptide chemistry. Removal of the group can be achieved by either hydrogenolysis or acidolysis (figure 1.22).

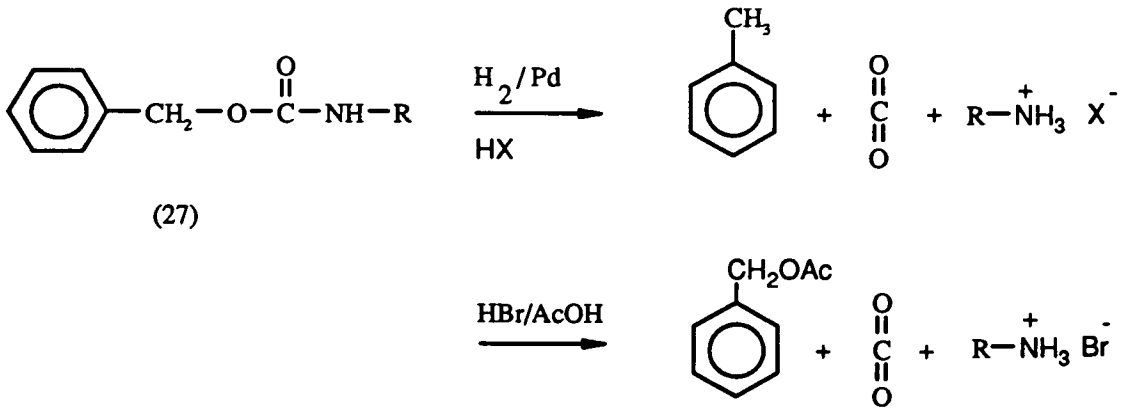


Figure 1.22 Removal of the Z Group

A demand for urethane protecting groups displaced under milder acidic conditions, led to the advent of tertiary butyloxycarbonyl (Boc, (**28**)) derivatives [170]. Urethanes of this type generate a very stable t-butyl carbocation (**29**) upon treatment of mild acid. Even greater acid lability is associated with the 2-(4-biphenyl)propyl(2)oxycarbonyl (Bpoc) derivatives (**30**) [171].

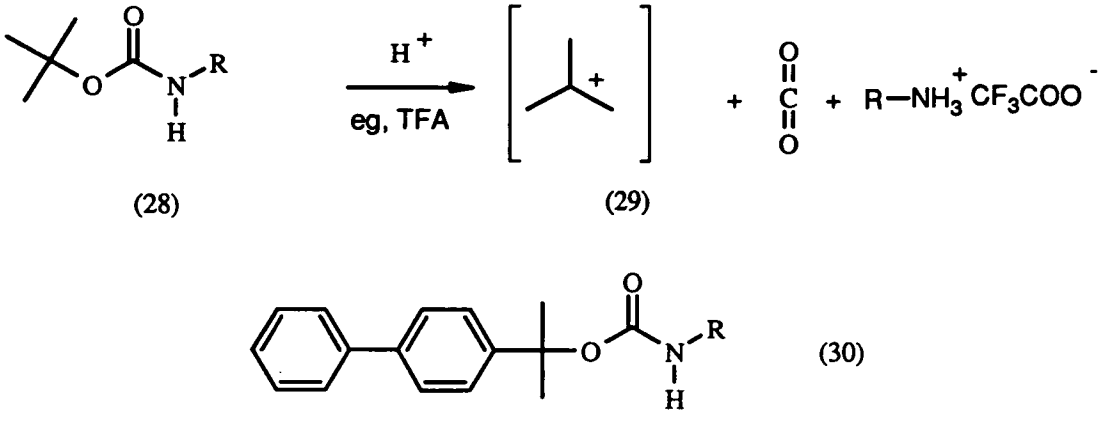


Figure 1.23 Acid Labile Urethanes

The difficulties associated with orthogonal protection strategies were greatly eased by the introduction of the base labile 9-fluorenylmethoxycarbonyl (Fmoc) urethane protecting group [172]. Deprotection of Fmoc-amino acids (**30**) is brought about by treatment with a secondary amine. This abstracts the proton on the fluorenyl sp^3

fluorenyl sp^3 carbon, generating a stable anionic species (**31**) which then undergoes β-elimination to liberate the α-amino group (see figure 1.24).

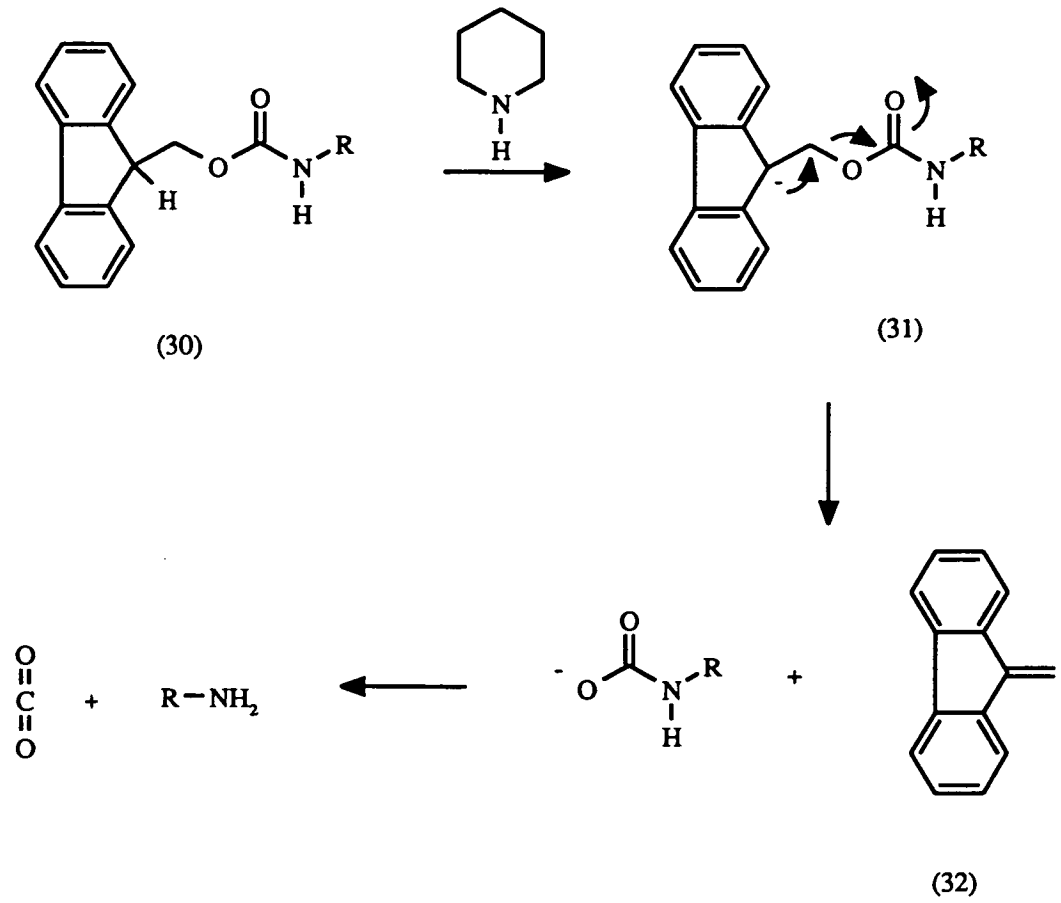


Figure 1.24 Deprotection of an Fmoc Group

The attraction of the Fmoc group is that it offers truly orthogonal peptide synthesis, since the side-chains of amino acids can be protected with acid labile groups.

1.3.5 Side-Chain Protecting Groups

In order to prevent branching of the growing peptide chain, as well as other well documented side reactions, it is necessary to protect the side chains of many amino acids [159,173]. The prerequisite for orthogonality means that the nature of this semi-permanent side chain protection will depend upon the choice of transient α-amino protection. Table 1.4 shows typical side-chain protecting groups used with the Fmoc and Boc N^{α} protecting groups.

AMINO ACID	FMOC STRATEGY	BOC STRATEGY
Lysine	Boc	2Cl-Z [174]
Arginine	Pmc [175]	NO_2, Tos
Serine, Threonine	Bu^t	Bzl
Aspartic & Glutamic Acid	OBu^t	OcHx [176]
Histidine	π-Bum [177], τ-Trt [178]	π-Bom [179], τ-Dnp
Cysteine	Acm [180], Trt, S-Bu^t	Acm, Bu^t, S-Bu^t
Tyrosine	Bu^t	Bzl
Glutamine, Asparagine	(Mbh) [181], Trt [182]	(Trt)
Tryptophan	None	Formyl
Methionine	None	None
Glycine	None	None
Alanine	None	None
Valine	None	None
Isoleucine	None	None
Leucine	None	None
Phenylalanine	None	None
Proline	None	None

( ) = optional

Table 1.4 Fmoc and Boc Protection Strategies

1.3.6 Synthetic Strategies

If we consider the construction of an imaginary tetrapeptide, ABCD, then clearly there are three possible synthetic strategies (see figure 1.25).

Assuming a similar yield for each coupling step, then route (ii) (fragment condensation) will always give a higher overall yield than either route (i) or (iii) (linear condensation). Indeed, with increasing length of peptide this advantage is enhanced.

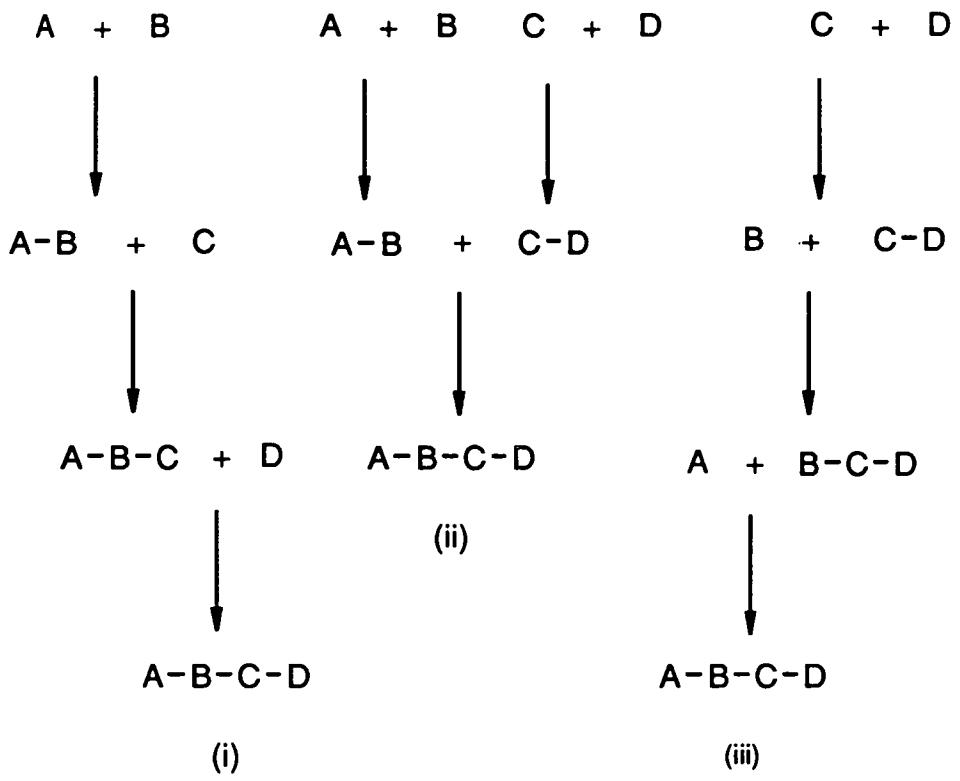


Figure 1.25 Three Approaches to a Tetrapeptide

Unfortunately, this simplistic view of things fails to take account of the chemistry involved. Routes (i) and (ii) both involve, at one stage or another, the activation of the carboxy-terminus of a peptide fragment, whereas route (iii) involves the acylation of an N^{α} peptide, except at the start. As we have already seen (section 1.3.4), activation of N^{α}-acylated amino acids or in this case a peptide, leads to 5(4H)oxazolone (**24**) formation and ultimately racemisation. For this reason, strategy (iii) (C - N linear condensation) is the most general approach to peptide construction. If the C-terminal amino acid of the peptide to be activated is glycine or proline, then clearly racemisation cannot take place and thus fragment condensation can be applied. However, this approach is limited to peptides containing these residues in a convenient position within the sequence.

1.3.7 Solution Phase Peptide Chemistry

Solution phase peptide chemistry has been successfully applied to the synthesis of many small peptides, such as oxytocin [183] and secretin [184]. In most cases, considerations of racemisation have led to the C to N linear condensation approach being adopted. In order to maintain orthogonality, the C-terminal carboxy-group

must be protected in an analogous manner to side-chain carboxyl groups. Figure 1.26 illustrates a suitable strategy for the synthesis of the tripeptide H-Lys-Asp-Ala-OH (33).

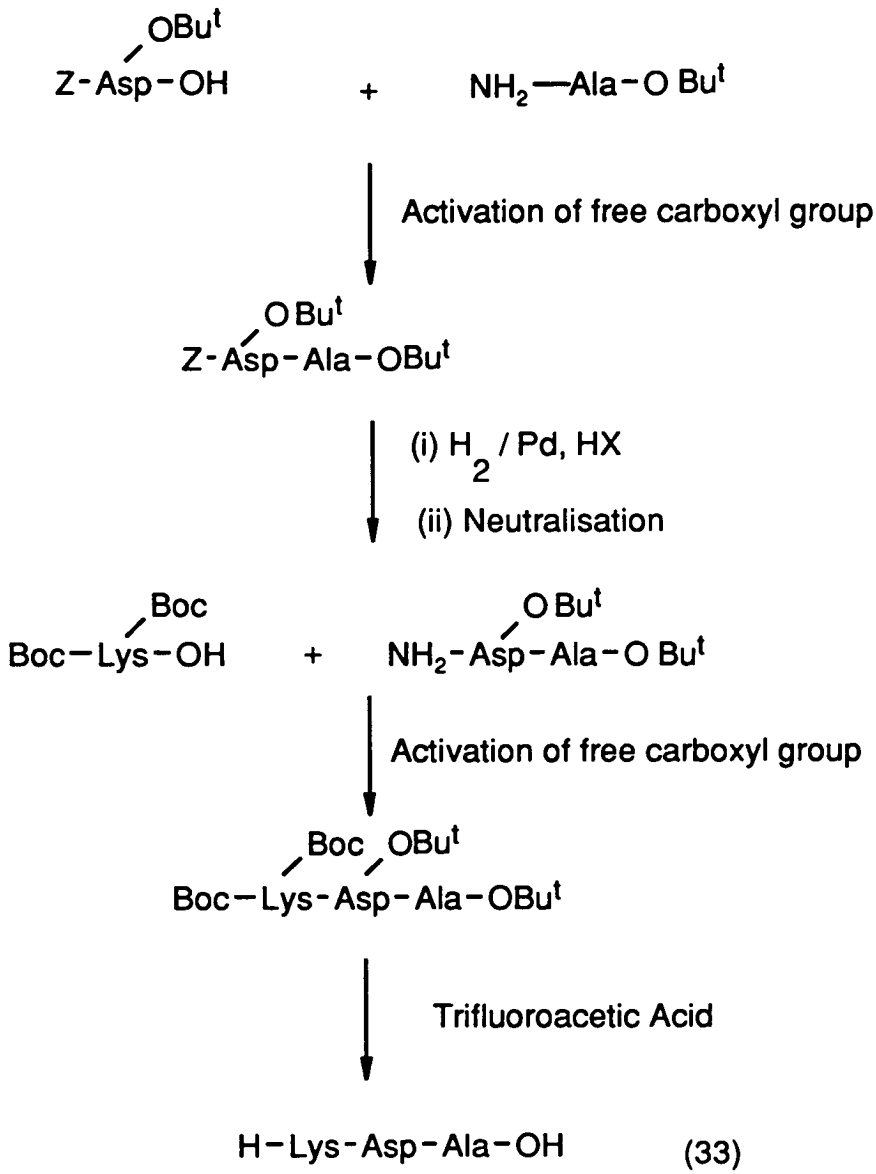


Figure 1.26 Solution Phase Construction of a Tripeptide

The four main problems inherent in solution phase synthesis are:

1. Excess activated amino acids required to optimize each coupling step must be removed from the reaction mixture.

2. Peptide intermediates have to be individually purified.

3. The number of acylation, deprotection and purification steps involved in constructing even the smallest of peptides, means all solution syntheses are extremely laborious

4. Protected peptides become increasingly insoluble with increased length, this severely restricts the application of the solution phase method.

1.3.8 Solid Phase Peptide Synthesis (SPPS)

In 1963, aware of the problems of peptide construction in solution, Merrifield introduced the concept of solid phase peptide synthesis [185]. His ingenious idea was to use an insoluble polymer support (a "resin") as a carboxy-terminal protecting group. Reaction of the C-terminal residue with a chloromethylated styrene-(1%)divinylbenzene copolymer (**34**), results in the attachment of the amino acid to the support (**35**). Elongation of the peptide chain can then be carried out using standard solution phase strategies (see figure 1.28). Note that growth of the peptide chain actually occurs within polymer beads, which are swollen in solvents such as N,N dimethylformamide (DMF).

Although Merrifield's original solid phase chemistry is still extensively used today, it does have one or two minor drawbacks. For example, the methodology is not strictly speaking orthogonal, both N^{α} and side-chain protecting groups are removed by acid, TFA and HF respectively. This results in a slow leakage of side-chain deprotection throughout the synthesis. Liquid hydrofluoric acid (HF) is an extremely corrosive acid and is not only difficult to handle, but can also catalyse several disturbing side-reactions such as N - O acyl-migration at serine residues (see figure 1.27).

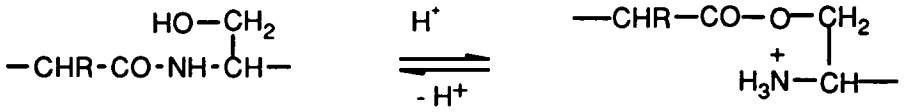


Figure 1.27 The N - O Acyl Migration

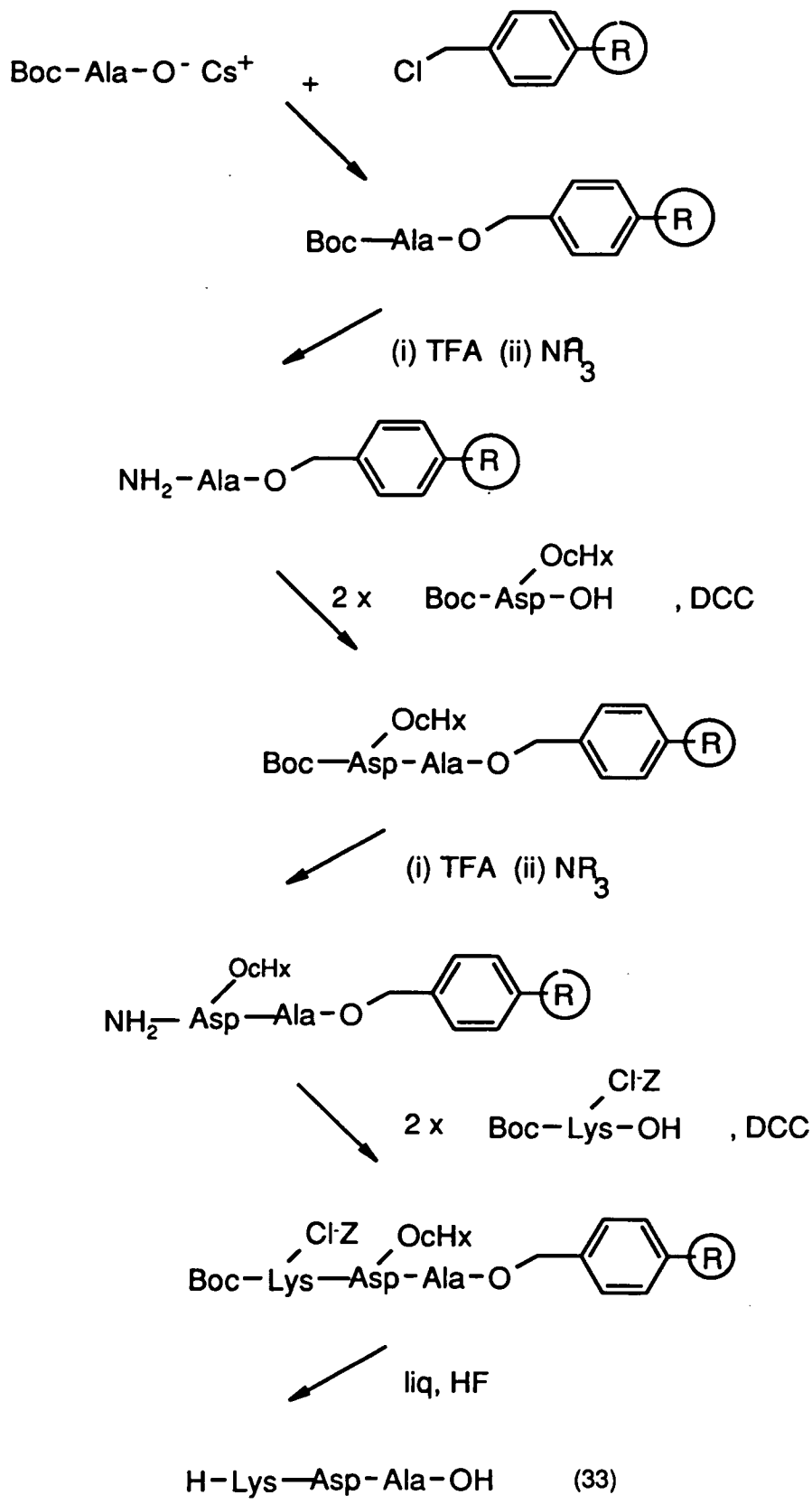


Figure 1.28 Merrifield SPPS of a Tripeptide

An alternative to the Boc-chemistry based Merrifield solid phase methodology uses Fmoc N^{α} protection and, hence, offers a means to truly orthogonal peptide synthesis (see figure 1.29). Furthermore, the use of acid labile side-chain protecting groups, as well as the incorporation of the acid sensitive *p*-alkoxybenzyl alcohol linking group [186] (Wang linker, **36**) between the peptide and the resin, means that final cleavage and deprotection can be accomplished under relatively mild conditions. This minimizes the undesirable side-reactions that occur upon cleavage of the peptide from the solid support.

SPPS has now developed into the general method for peptide synthesis as it has the following advantages over the solution method.

1. Simplified procedures: - unreacted reagents and by-products can be simply washed away.

2. No need to isolate intermediates, therefore reduction in mechanical losses in addition to quicker syntheses.

3. The methodology may be automated.

Solid phase peptide synthesis offers a reliable and efficient approach to the construction of peptides, including those containing unnatural amino acids.

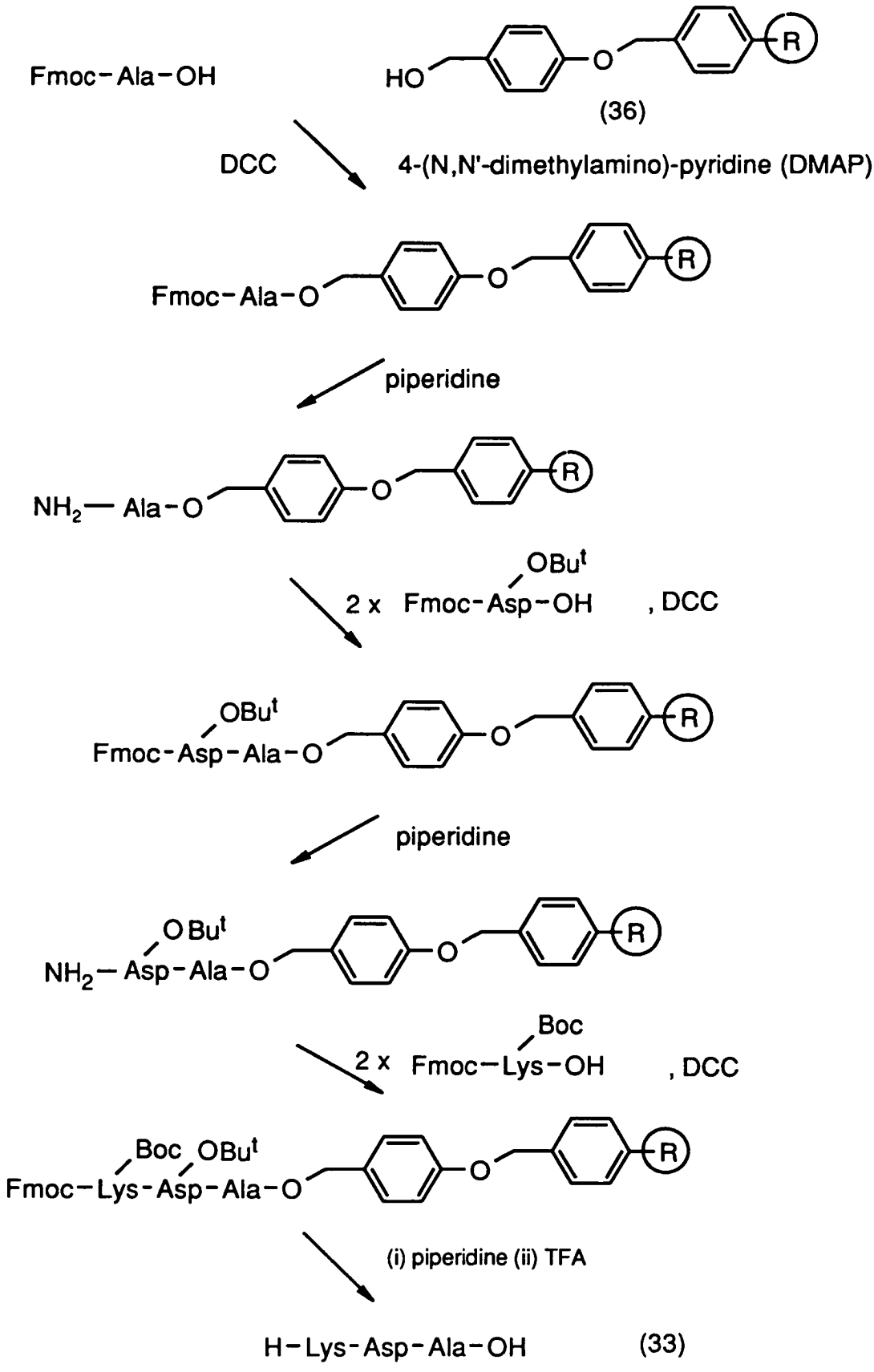


Figure 1.29 Modified SPPS of a Tripeptide

Chapter 2 - Discussion

Section 2.1 Aims of the Project

Solid phase peptide synthesis offers access to protein and peptide analogues inaccessible via recombinant techniques. Such derivatives could conceivably facilitate easier study of complex processes like folding. However, before these analogues can be chemically synthesised, it is necessary to prove beyond any doubt, that any artificially constructed wild type protein behaves in an identical manner to its cellular counterpart. Only then will any subsequent work have relevance.

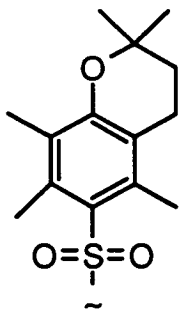
We chose to investigate the structural stability of ubiquitin by the construction of specific analogues. To begin with though, we were fully aware of the necessity of comparing the conformation of a synthetically constructed ubiquitin with that of a ribosomally constructed ubiquitin.

At the outset of this work, the synthesis and purification of a polypeptide the size of ubiquitin was certainly at the very limit of what could be achieved by linear solid phase peptide synthesis. Therefore, our initial goal was to optimize both the synthesis and the purification of the protein, so as to facilitate the isolation of a homogeneous material. Having isolated and chemically characterized the synthetic ubiquitin, it was then our intention to study its structure and directly compare this with that of the natural material. Furthermore, it was necessary to assay the biological activity of synthetic ubiquitin and make a quantitative comparison with its natural counterpart.

Assuming our synthetic ubiquitin resembled that of the natural material, it was planned to construct and study the behaviour of a carefully chosen specific ubiquitin analogue.

Section 2.2 Background to Present Work

An early attempt in this laboratory at the chemical synthesis of ubiquitin was reported in 1989 by Ramage and co-workers [187]. Synthesis of the polypeptide was carried out using an N^{α}-Fmoc SPPS strategy in which the side-chains of arginine residues were protected with the newly developed 2,2,5,7,8- pentamethylchroman-6-sulphonyl (Pmc) group (**37**).



(37)

Figure 2.1 The Pmc Protecting Group for Arginine

An in house approach to real time assessment of the progress of the synthesis (see section 2.4 for details of monitoring system) had revealed several problematic stages during the construction of the ubiquitin sequence. It was therefore no surprise that purification of the required material from the crude cleavage mixture proved difficult. The multi-step purification protocol eventually used comprised gel filtration chromatography, various types of ion exchange chromatography, chromatofocussing and semi-preparative high performance liquid chromatography (HPLC). However, this system was by no means ideal. For example, ion exchange chromatography was nowhere near as successful as had been hoped, probably because the conditions used were not optimum for ubiquitin. Nevertheless, this early work did lay the foundations for the more fruitful efforts which followed.

A second synthesis of ubiquitin carried out by Briand and co-workers[188], fared little better than the original work of Ramage. This group had elected to use Boc chemistry with SPPS. Subsequent purification by gel filtration, medium pressure liquid chromatography and ion exchange chromatography afforded only 7mg of homogeneous synthetic ubiquitin. However, this group were able to show that this material was immunoreactive towards anti-ubiquitin antibodies and in addition demonstrated a conjugation activity comparable with natural ubiquitin.

Section 2.2 Purification of Bovine Ubiquitin

Given the apparent inadequacies of existing purification protocols, our initial priority was to develop a more effective approach to the purification of synthetic ubiquitin. Commercially available bovine ubiquitin (Sigma) is not only comparatively cheap, but more importantly, surprisingly impure (approx 80%).

Therefore it seemed rational to take advantage of this lack of homogeneity whilst designing a new purification strategy and remove the contaminants in the commercial sample. Such a protocol could then hopefully be successfully applied to the purification of synthetic ubiquitin, although of course the contaminants will be different.

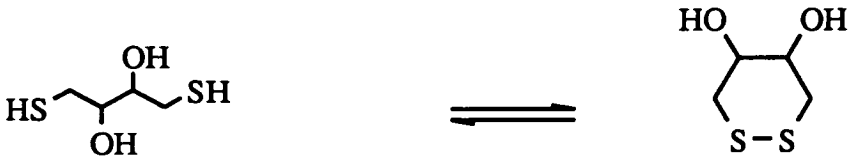
Two features of the original mode of purification, gel filtration chromatography and dialysis, were deemed sufficiently useful to be retained in the new strategy. The remainder of the steps were devised independently of any previous work. Crude bovine ubiquitin was subjected to the four step purification outlined below.

1. Gel Filtration Chromatography

This relies on the principle of size exclusion to separate the components of a mixture. The filtration matrix consists of swollen porous beads which can absorb molecules up to a certain size. Therefore, the small and large molecules that comprise a mixture will be divided into an absorbed fraction (stationary phase) and an unabsorbed fraction (mobile phase) respectively. If a mixture is pumped through a column of gel filtration beads then this partitioning will result in the larger molecules of the mixture being eluted before the smaller ones.

Of the various types of gel filtration media commercially available, we chose to use Sephadex G50 (fine) beads (Pharmacia). This packing is particularly useful for separations in the 500-10000 Da molecular weight range in which we are operating. Crude bovine ubiquitin was dissolved in a running buffer containing 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5. The components of this buffer were chosen carefully and with particular attention to the properties of crude synthetic ubiquitin. High concentrations of urea are known to solubilise protein suspensions. While this is of little consequence to the highly soluble bovine ubiquitin, the crude synthetic material obtained immediately after cleavage was known to be very difficult to get into solution. In addition, 8 M urea is known to denature ubiquitin[112], thereby conveniently erasing any of the misfolded synthetic ubiquitin molecules which might be expected to form during the harsh cleavage conditions. The sacrificial reducing agent dithiothreitol (DTT, **38**) was included in the hope of minimizing the easier methionine oxidation that is associated with the unfolded state [136].

Under these conditions gel filtration chromatography was found to remove the small contaminants found in crude bovine ubiquitin. Crude synthetic ubiquitin would of course contain small truncated peptide impurities, as well as organic scavenger molecules left behind after initial cleavage off the polymer support. All of these, it was hoped, could be resolved from the remainder of the mixture using this technique.



(38) Reduced (38) Oxidised

Figure 2.2 Redox equilibrium of DTT

2. Dialysis

Dialysis is a purification technique which, like gel filtration, separates the components of a mixture based on molecular size and to a lessor extent charge. A typical dialysis membrane consists of amorphous regenerated cellulose in which the individual chains are crosslinked in such a way as to create pores. The degree of crosslinking determines the size of these pores through which sufficiently small molecules can pass. Dialysis membranes of various porosity are commercially available, usually in the form of a cylindrical bag. A dialysis bag with a molecular cut-off of approximately 2000 Da was chosen for this work. If a ubiquitin solution is sealed in one of these bags surrounded by an aqueous solution, then anything in the protein solution of molecular weight less than about 2000 will be free to permeate through the bag until its concentration is the same either side of the membrane (see figure 2.3). The concentration of solute molecules left in the bag will of course be negligible if a large volume of external solution is used.

Dialysis was principally used as a means by which to control the refolding of our denatured post gel filtration ubiquitin. Slowly removing the denaturant from the internal solution, by dialysing against ever decreasing external concentrations of urea and DTT, should allow the protein to gradually adopt its native conformation.

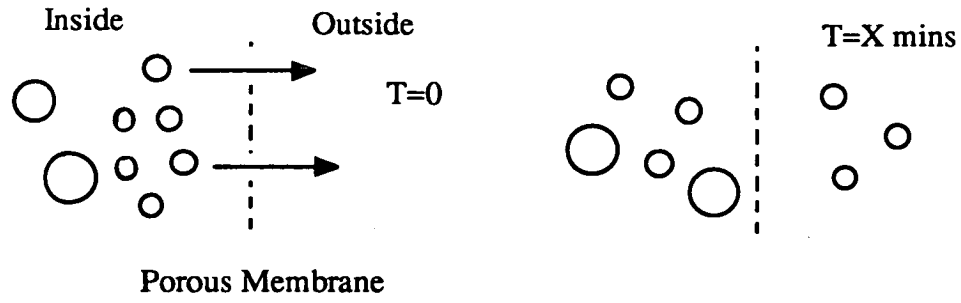


Figure 2.3 Principle of Dialysis

3. Ion Exchange Chromatography

Gel filtration and dialysis are convenient techniques for the removal of contaminants which differ markedly in size from the required material. However, removal of similarly sized impurities requires a different set of purification criteria. Ion exchange chromatography satisfies this requirement since it relies on ionic charge to separate a mixture.

Peptides and proteins usually possess a nett positive or negative overall charge. This arises from the combination of individual charges on the basic (+ve) and acidic (-ve) side-chains of some amino acids (principally Glu, Asp, Lys & Arg). The individual pKa's of these residues results in the value and even the sign of the overall charge on a protein varying with pH. Furthermore, every protein has an associated pH, known as the isoelectric point (pI), at which the total positive charge exactly cancels out the total negative charge. If the pH is raised above the pI then the protein will progressively become more negatively charged, whereas the further the pH drops below the pI then the more positively charged it will become. Ion exchange exploits the fact that the components of a protein mixture each have a different pI and so at any given pH will bind more or less tightly to a counter charged ion exchange matrix. The binding affinity is also inversely related to the ionic strength (Debye-Huckel) hence a combination of changing the pH and ionic strength, carried out by adjusting the concentration of the buffer, can be tailored to the separation in question.

a) Cation Exchange Chromatography

In cation exchange the immobilised counter-ion on the exchanger must be negatively charged. We decided to use the carboxy-methyl based exchanger, CM-Sepharose. Further purifying bovine ubiquitin by cation exchange requires the use

of a pH on the acidic side of its isoelectric point (pI=6.7). Therefore, semi-crude bovine ubiquitin was loaded onto the cation exchange column in a buffer containing 50 mM NH_4OAc at pH 4.5. An initial pH gradient going from pH 4.5 to pH 5.5 was found to elute various small contaminants, however the bulk of the protein remained bound to the column. A subsequent salt gradient (50 mM to 0.3 M NH_4OAc at pH 5.5) did elute the material from the column, with good resolution being achieved between the different components of the mixture.

b) Anion Exchange Chromatography

The fourth and final step in the purification was anion exchange chromatography which was necessarily carried out at a pH above the pI of ubiquitin. This technique requires the immobilised counter-ion to be positively charged and to this end the diethylaminoethane (DEAE)-Sepharose exchanger was chosen. Bovine ubiquitin having already been substantially purified by the above, was applied to the column in a buffer containing 50 mM NH_4HCO_3 at pH 9.3. Again a salt gradient (running from 50 mM to 0.3 M NH_4HCO_3 at pH 9.3) was found to give essentially baseline resolution of the eluted components, by far the major of which was the required material (see figure 2.4). Anion exchange proved to be exceptionally sensitive to pH and even a minor deviation from the optimum pH of 9.3 resulted in an appreciable decrease in resolution.

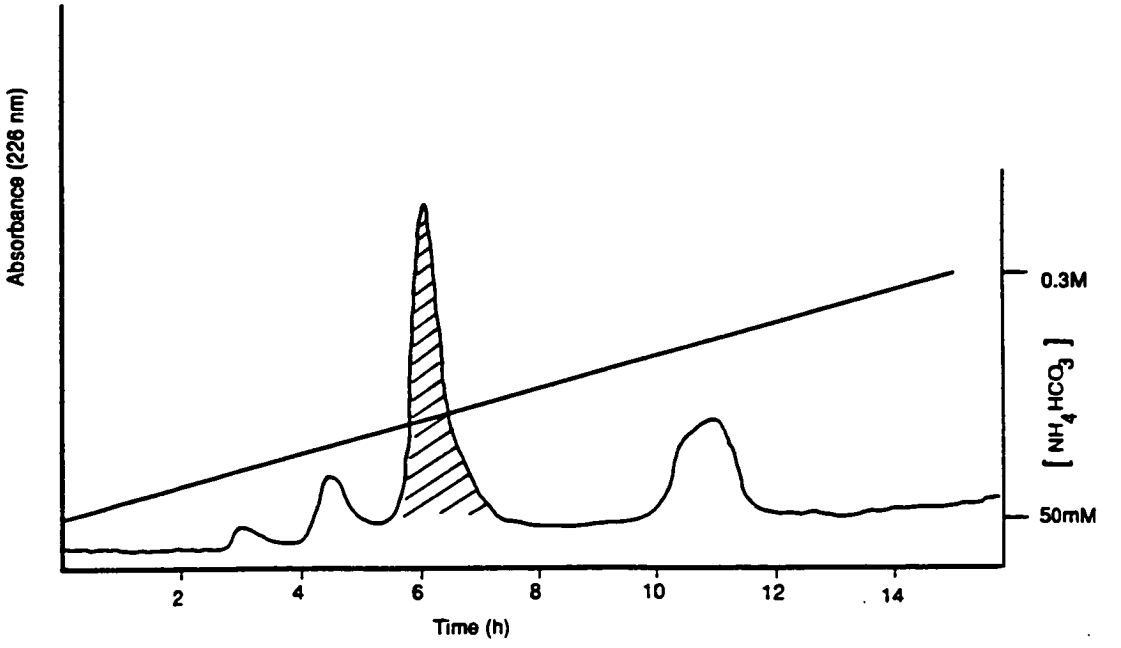


Figure 2.4 Anion Exchange Chromatography on Bovine Ubiquitin

Analysis of purified bovine ubiquitin by isoelectric focussing electrophoresis (IEF) and analytical HPLC suggested its homogeneity had risen from around 80% to well in excess of 95%. The success of this protocol in purifying bovine ubiquitin was beneficial in two ways:

(1) It offered a new approach to the purification of a synthetic ubiquitin.

(2) It had afforded us with an exceptionally pure natural material which could be more easily used in comparative studies with a synthetic ubiquitin.

Section 2.4 Synthesis and Purification of Ubiquitin

2.4.1 SPPS of Ubiquitin

Our approach to the solid phase synthesis of ubiquitin employed the N^{α}-Fmoc strategy introduced by Carpino [172] and subsequently developed by Shepard [189]. The invaluable information gleaned from previous attempts at the synthesis of ubiquitin in Edinburgh [187] was assimilated into our revised synthetic strategy. Constructing ubiquitin from the C-terminal towards the N-terminal requires the initial attachment of Gly-76 to the polymer support. Conversion of Fmoc-Gly-OH to the amino acid chloride (**39**), by treatment with thionyl chloride, allowed the *p*-alkoxybenzyl alcohol resin [186] to be functionalized with this N^{α}-Fmoc protected amino acid (see figure 2.5).

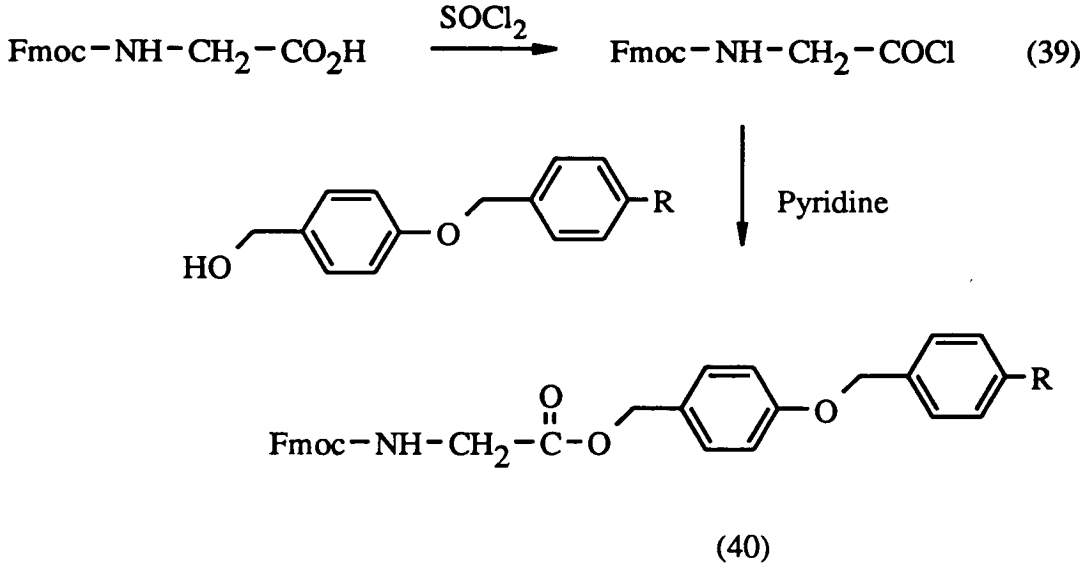
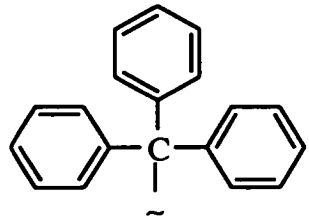


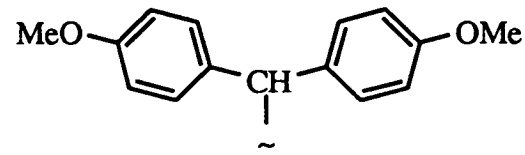
Figure 2.5 Loading of the Resin

Of special concern at this stage was the number of available sites on the resin actually acylated with the Fmoc protected amino acid. Experience had suggested that a resin with high initial loading can lead to steric interactions occurring between growing peptide chains, especially towards the end of long syntheses. Such interactions could conceivably reduce yields in coupling steps, thereby accounting for some of the problems encountered in previous syntheses in which the degree of loading was as high as 70% of the available sites (0.44 mmol of glycine per gram of resin). It was therefore decided to substantially reduce the functionality of the resin in an effort to minimize these unfavourable interactions. This was ultimately achieved by reducing the reaction time to 1 hour and molar ratio of Fmoc-Gly-Cl to resin to 2:1. Under these conditions the level of loading was lowered to around 40% (0.28 mmole per gram) as established by UV analysis of a loaded resin sample deprotected by 20% piperidine in DMF. In addition, no glycylglycine content could be detected by amino acid analysis, indicating the shorter reaction time had eliminated the small amount of polymerisation often associated with loading of glycine. The remaining 60% of functional groups on the resin were capped with benzoylchloride so as to eliminate the possible build up C-terminal deletion peptides during the synthesis.

Where necessary, amino acid side-chains were protected with the standard acid labile groups used with Fmoc based SPPS (see table 1.4). For example, the guanidino group of arginine was protected by the Pmc group and the τ-nitogen in the imidazole ring of histidine with the trityl (Trt) group (**41**). In a departure from previous syntheses, the amide groups of Asn and Gln were protected with the 4,4'-dimethoxybenzhydryl (Mbh) group (**42**). These two amino acids had previously been left unprotected.



(41)



(42)

Figure 2.6 The Trityl and Dimethoxybenzhydryl Protecting Groups

Solid phase peptide synthesis lends itself to automation and this synthesis of ubiquitin was carried out on an Applied Biosystems Model 430A peptide synthesiser. In an effort to obtain near quantitative acylation at each stage of the synthesis most of the amino acids were coupled once as their symmetrical anhydride and then twice as their HOBt active esters, each time using DIC as the condensing agent. The exceptions to this were Fmoc-His(Trt), Fmoc-Asn & Gln(Mbh) and Fmoc-Gly. Activation of Fmoc-His(Trt) as the symmetrical anhydride is known to lead to racemisation via the mechanism shown in figure 2.7. Prevention of this can be achieved by using a slightly less active species (rendering the carbonyl group less electrophilic) during the coupling step [190]. To that end Fmoc-His(Trt) was coupled three times as the HOBt active ester.

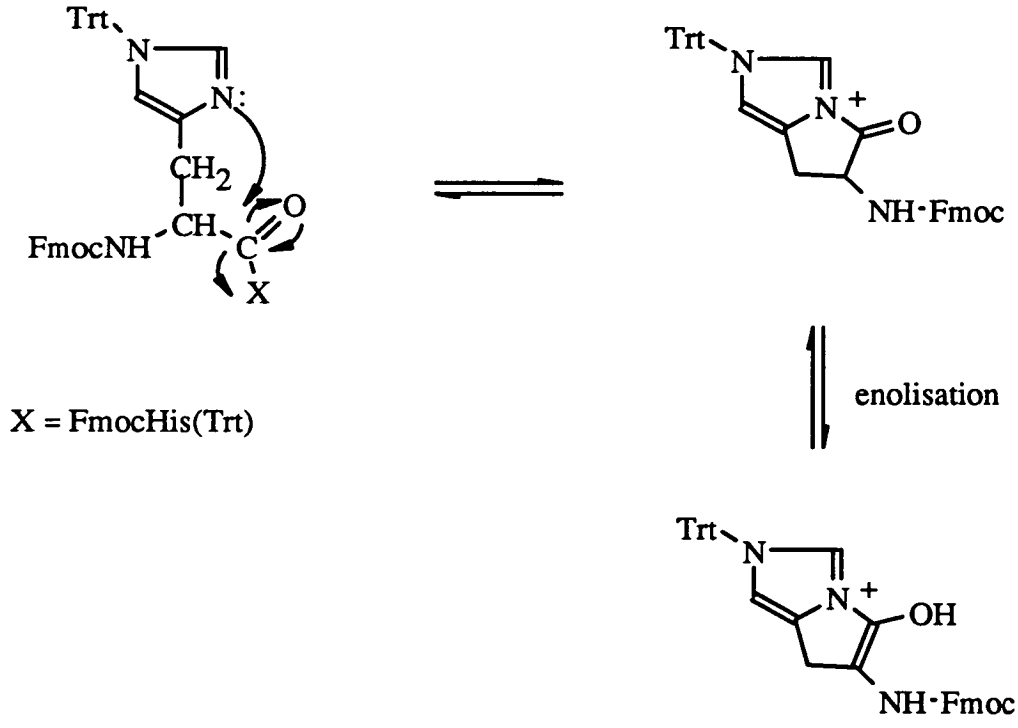
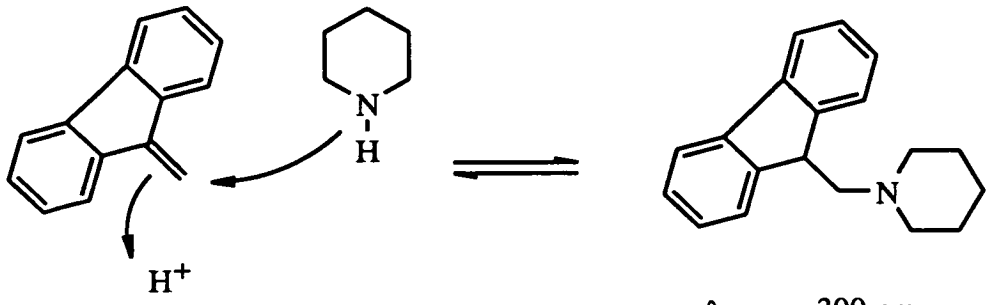


Figure 2.7 Racemisation of Fmoc-His(Trt) Anhydride

Considerations of cost meant that Fmoc-Asn(Mbh) and Fmoc-Gln(Mbh) were also coupled three times as their HOBt active esters. Finally, Fmoc-Gly was coupled once in a single symmetrical anhydride step so as to avoid formation of glycylglycine which can form as a result of multiple coupling procedures. Every acylation in the synthesis without exception was carried out using a 1:1 mixture of

DMF:dioxan as the solvent, previous syntheses had used 100% DMF. Inclusion of the dioxan was observed to give the peptide-resin better swelling properties and hence render the growing peptide more accessible to peptide synthesis reagents[191].

Immediately after each coupling, acetic anhydride in DMF was added to cap any unreacted N^{α}-amino groups remaining on the peptide-resin, thus preventing the accumulation of deletion peptides. Capping was followed by deprotection of the N^{α}-Fmoc group thereby exposing the next amino group to be acylated. Fmoc deprotection involved successive treatments with 20% piperidine in DMF for periods of 5, 3, 3 & 1 min. In order for real time assessment of the synthesis a diluted aliquot of each deprotection solution, containing the chromophoric fulvene-piperidine adduct (**44**), was monitored in a continuous flow mode from the synthesiser to a UV detector set at 302nm. An integrator allows the size of every deprotection peak to be compared with that of the previous amino acid. In this way the relative efficiency of each coupling and deprotection step can be monitored and the potential yield of the protein upon completion of the synthesis can be gauged. Moreover, unlike ninhydrin analysis [192] this approach does not require the removal of any resin from the reaction vessel thus maximising the final protein yield.



(43) (44)

Figure 2.8 Fulvene-Piperidine Adduct Formation

On completing the assembly of the protected ubiquitin the final weight of peptide-resin was 3.01 g compared to the initial polystyrene weight of 1.15 g. Figure 2.9 illustrates the deprotection profile obtained for the ubiquitin synthesis from which the expected yield of peptide on the resin was approximately 70%. This

corresponds to an average yield of 99.8% for each of the 75 coupling steps and 75 deprotection steps in the synthesis.

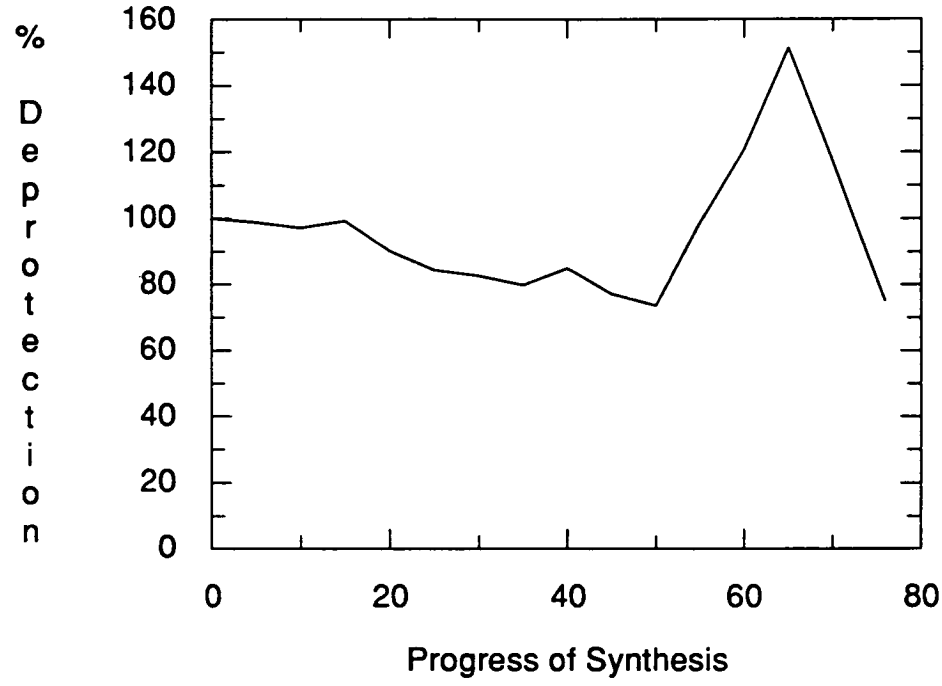


Figure 2.9 Deprotection Profile for the Ubiquitin Synthesis

The final amino acid, methionine, was left with its Fmoc group still attached so as to allow a quantitative assessment of the amount of this residue linked to the peptide. Deprotection of two samples of Fmoc-peptidyl resin with 20% piperidine in DMF followed by UV analysis of the piperidine adduct (**44**) gave yields of 48% and 58% for the final Fmoc-Met loading relative to the initial Fmoc content for Gly-76. These values probably represent a more realistic estimate of the yield on the resin than the rather erratic results obtained from on-line monitoring towards the end of the synthesis.

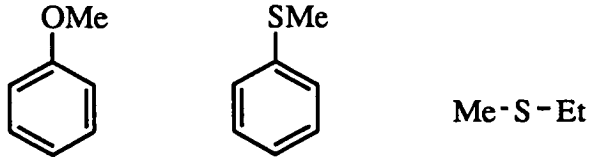
2.4.2 Purification and Characterization of Synthetic Ubiquitin

The large quantity (3g) of peptide-resin obtained from the synthesis meant that the purification had to be carried out in three batches. Note that all weights referred to in this section are in effect aggregate values.

Before any purification could be carried out it was first necessary to cleave the polypeptide from the polymeric support and at the same time remove all side–chain

protecting groups. The N^{α}-Fmoc protecting group on the peptidyl resin was initially removed with 20% piperidine in DMF. To remove the side-chain protecting groups and to cleave the peptide from the support, the peptidyl resin was treated with a mixture of thioanisole (5%, v:v), ethyl methyl sulphide (5%) and anisole (5%) for 20 minutes under nitrogen before adding water (5%) and trifluoroacetic acid (80%) and stirring at room temperature for 3.5 hours.

Anisole (**45**) and thioanisole (**46**) were added to scavenge the Bu^{t}, Mbh and Trityl carbocations produced as by-products of the cleavage, thereby preventing alkylation of tyrosine and phenylalanine residues in the peptide. The anti-oxidant ethyl methyl sulphide (**47**) was added as a safeguard against methionine oxidation.



(45) (46) (47)

Figure 2.10 Scavengers used in Cleavage Mixture

The optimum cleavage time was established by conducting a small scale (50 mg) experiment using the same concentrations of cleavage reagents to the above, but in reduced volumes. Samples of the TFA mixture were periodically removed from this trial cleavage and triturated with diethyl ether to remove scavengers and to precipitate the peptide. HPLC analysis of the resulting crude mixtures dissolved in aqueous acetic acid revealed a period of 3.5 hours to be the most suitable (see figure 2.11).

Following deprotection and cleavage the TFA was removed *in vacuo* leaving a brown residue which was subsequently triturated with diethyl ether containing 2% β-mercaptoethanol. The precipitated peptide was immediately added to a nitrogen saturated gel filtration buffer containing 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5. This suspension was stirred for 16 hours to allow the crude peptide to fully dissolve.

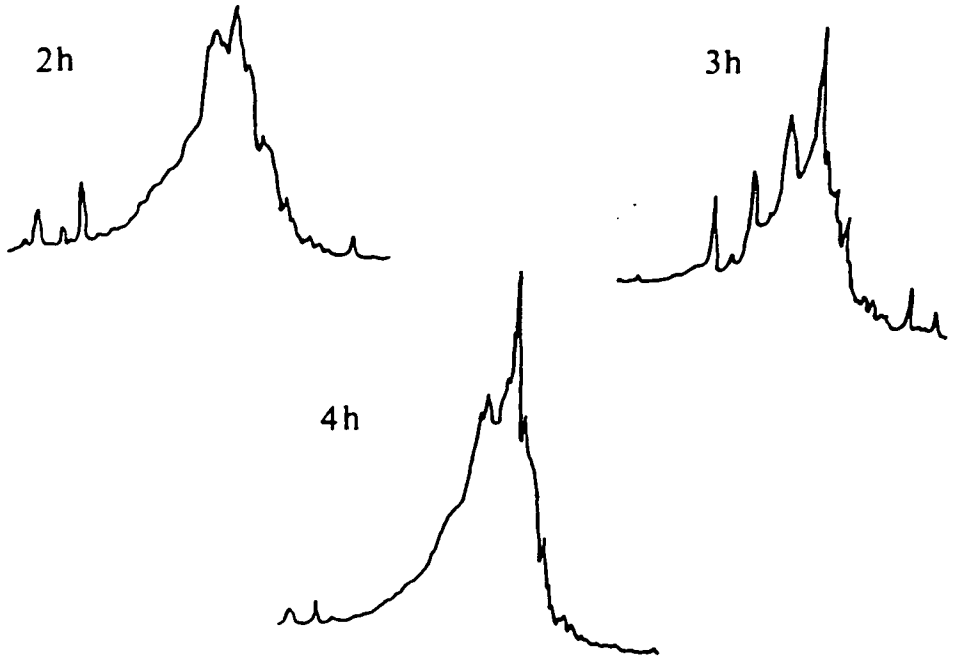


Figure 2.11 HPLC Traces obtained from Trial Cleavage (C_8 RP300, gradient system A)

Purification of the required sequence from the crude mixture was accomplished using the four step procedure previously developed using bovine ubiquitin (see section 2.3).

1. Gel Filtration Chromatography

The polypeptide solution was subjected to gel filtration on a Sephadex G50 (fine) column previously equilibrated with a nitrogen saturated buffer containing 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5 (see figure 2.12a). The required protein fractions, from which small organics and truncated peptides had been removed, were isolated and subsequently dialysed.

2. Stepwise Dialysis

The post gel filtration synthetic ubiquitin was allowed to renature by dialysing it against gradually decreasing concentrations of urea and dithiothreitol. On completion of the dialysis the solution was concentrated *in vacuo* to 20-30 ml in volume and a small sample analysed by HPLC (see inset to figure 2.12a) and amino acid analysis (table 2.1).

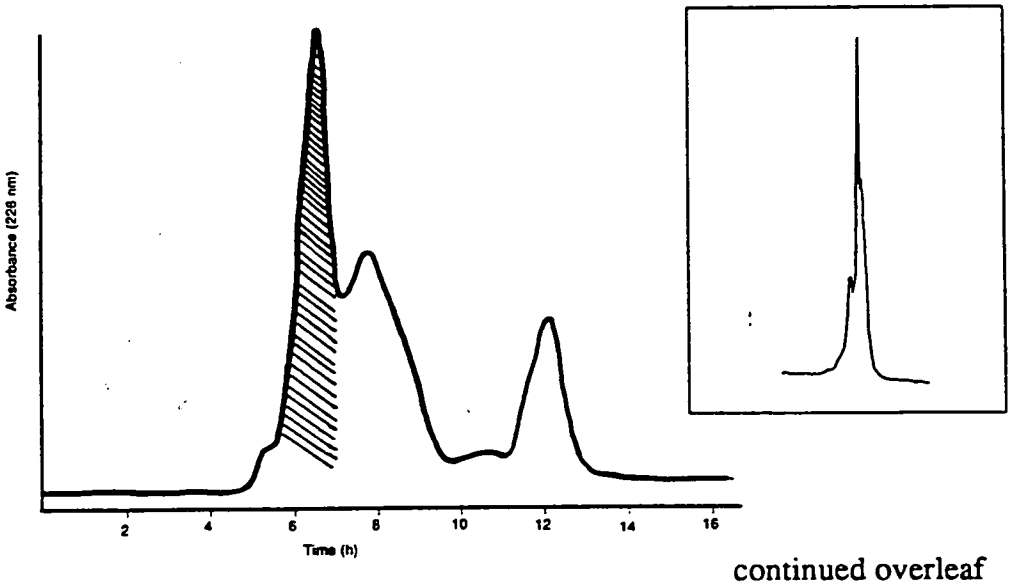
3. Cation Exchange Chromatography

The desired material from dialysis was further purified by cation exchange chromatography on a CM-Sepharose column pre-equilibrated with starting buffer. After loading the synthetic material onto the column an initial pH gradient running from pH 4.5 to pH 5.5 was found to elute some contaminants, however the bulk of the polypeptide remained bound to the column. Having been left to run isocratically at pH 5.5 for approximately 0.5 column volumes a salt gradient of 50 mM to 0.3 M NH_4OAc at pH 5.5 was applied to elute the protein from the column (figure 2.12b). The resulting fractions were isolated and combined on the basis of their HPLC spectra (inset to figure 2.12b) and elution times. Repeated lyophilisation was found to remove most of the salt from the combined fractions which were then analysed by amino acid analysis (table 2.1).

4. Anion Exchange Chromatography

The required fractions from the CM-Sepharose column were dissolved in 50 mM NH_4HCO_3 at pH 9.3 and loaded onto a pre-equilibrated DEAE-Sepharose column. A linear salt gradient of 50 mM to 0.3 M NH_4HCO_3 at pH 9.3 was applied and gave rise to two peaks, labelled A and B (see figure 2.12c). The fractions corresponding to these two peaks were isolated and repeatedly lyophilised to remove salt, thereafter affording 71.8 mg of A and 21.7 mg of B. As with the bovine ubiquitin purification, anion exchange chromatography proved to be remarkably sensitive to pH with pH 9.3 giving markedly better resolution over any surrounding pH.

a)



continued overleaf

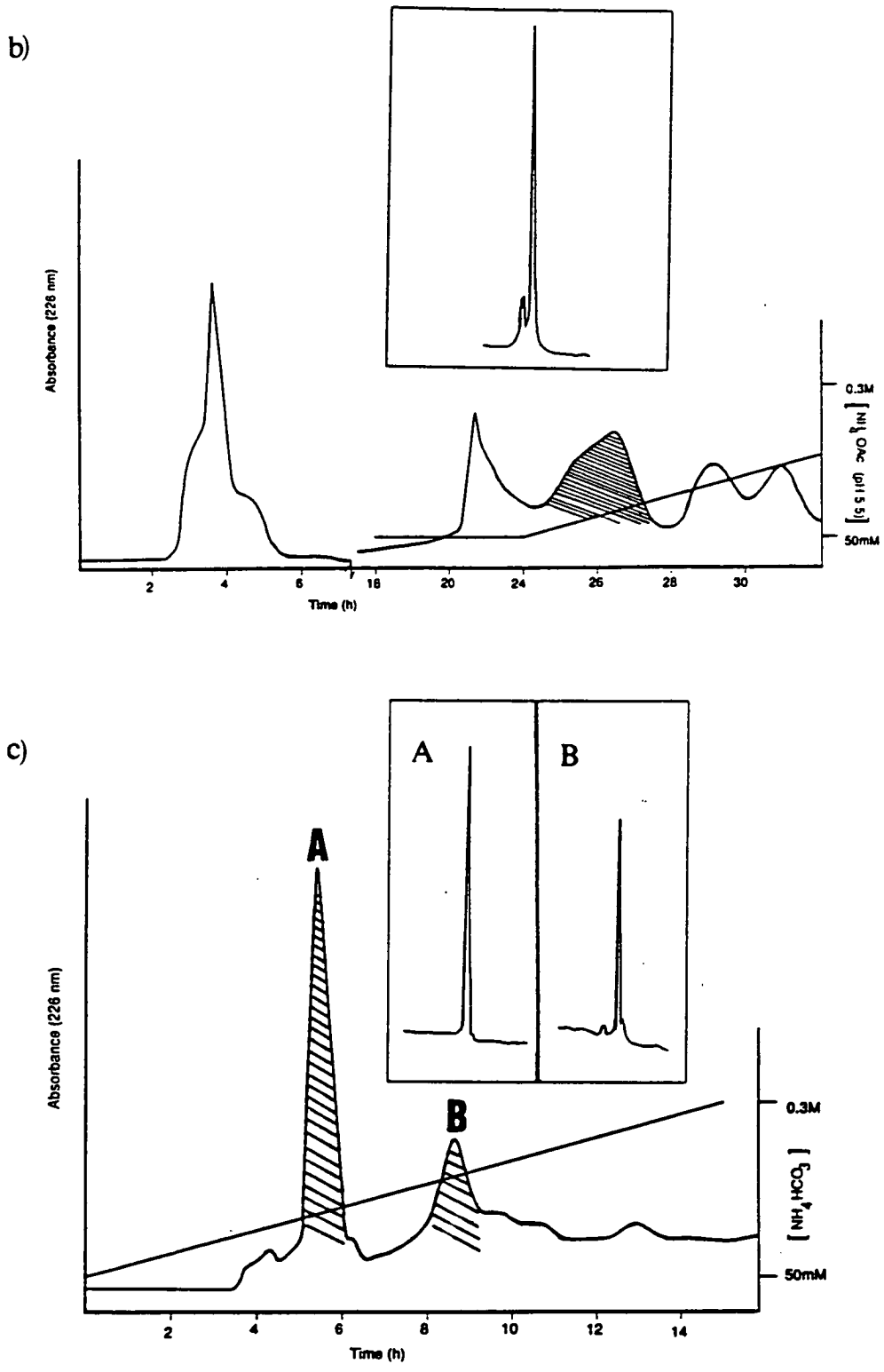


Figure 2.12 Purification of Synthetic Ubiquitin
a) Gel Filtration b) Cation Exchange c) Anion Exchange
Shaded areas are the protein fractions isolated. Insets show analytical HPLC spectra (C_8-RP300, gradient A) obtained.

Analysis of A & B by analytical HPLC (inset to figure 2.12c) revealed they both possessed single peak purity. Furthermore, both had identical retention times to bovine ubiquitin on the C_8 RP300 column (16.9 mins, gradient system A). Amino acid analysis further highlighted the similarities between synthetic A and B, both of which had amino acid compositions entirely consistent with ubiquitin (see table 2.1).

Amino Acid	Expected Values	Gel Filt. Dialysis	CM-Sepharose	DEAE Syn A	DEAE Syn B
Asx	7	7.44	7.40	7.54	7.34
Thr	7	5.12	4.12	5.88	5.36
Ser	3	2.64	0.78	2.29	1.85
Glx	12	10.82	13.62	13.00	12.50
Pro	3	-	3.30	3.54	2.29
Gly	6	5.66	5.83	6.29	5.87
Ala	2	1.98	1.79	2.05	1.90
Val	4	4.64	4.00	4.02	3.94
Met	1	0.39	0.95	0.95	1.00
Ile	7	5.83	6.53	6.78	6.73
Leu	9	8.11	9.29	9.22	9.10
Tyr	1	1.45	1.26	0.98	1.15
Phe	2	1.65	2.11	2.06	2.07
His	1	0.99	1.02	0.94	0.98
Lys	7	5.28	7.02	7.04	6.98
Arg	4	3.74	4.06	4.05	4.07

Table 2.1 Amino Acid Analysis at Various Stages of the Purification 40 h hydrolyses in 6 N HCl (5 mg/ml Na_2SO_3) at 110°C

The isoelectric points of synthetic A and B were determined by Isoelectric Focussing Electrophoresis (IEF) to be identical with that of purified bovine ubiquitin (pI=6.7). Moreover, IEF provided additional evidence for the high level of homogeneity initially suggested by HPLC and amino acid analysis (see figure 2.13).

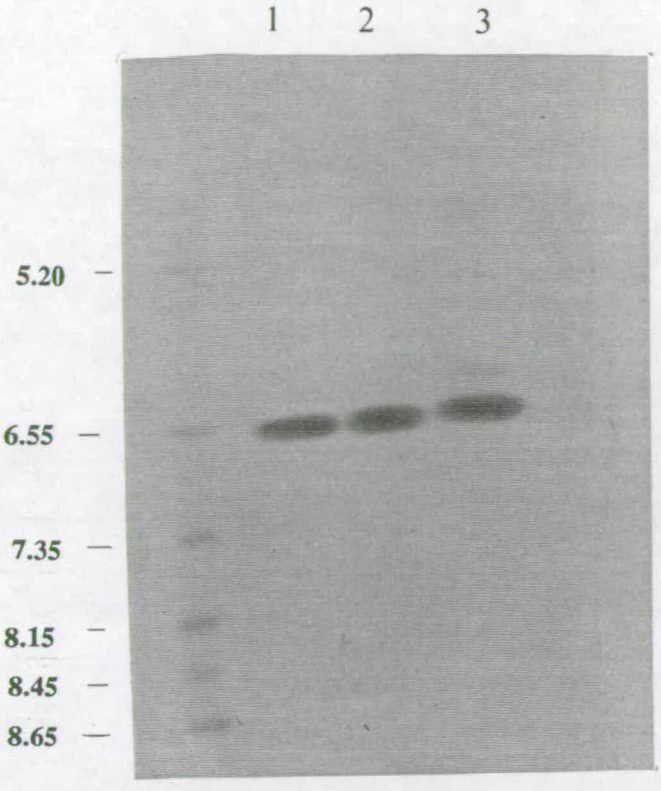


Figure 2.13 Isoelectric Focussing Electrophoresis
lane1 = bovine Ub, lane2 = synthetic Ub A
lane3 = synthetic Ub B

An antibody binding assay (Western Blot) was carried out to compare the immunoreactivity of A & B with purified bovine ubiquitin (see figure 2.14). After an initial IEF gel the proteins were electroblotted onto nitrocellulose and probed with rabbit anti-ubiquitin IgG polyclonal antibodies. An anti-rabbit IgG horse-raddish peroxidase conjugated secondary antibody was then used to visualize the initial anti-ubiquitin antibody. On staining this secondary antibody it was apparent that both synthetic materials had very similar immunoactivities to purified bovine ubiquitin. Although this is additional evidence for the equivalence of A, B and bovine ubiquitin, it should be noted that polyclonal antibodies would more than likely recognize continuous epitopes in a truncated or deleted ubiquitin peptide. For this reason such a result cannot be construed as a conclusive piece of evidence for the three materials being the same.

Figure 2.14 Western Blot of Bovine Ub (lane1), Synthetic A (lane 2) and Synthetic B (lane3).

A preliminary mass determination using Laser Desorption Time-Of-Flight Mass Spectrometry (LD-MS [193]) revealed A and B to have masses of 8568.5 and 8571 Da respectively. Given that the theoretical mass of ubiquitin is 8565 Da, both these experimentally obtained masses were within the error limitations of the instrument and hence both materials appeared to have the correct mass for ubiquitin.

Amino acid analysis and mass spectrometry together suggested synthetic A and B had the correct chemical composition for ubiquitin. However, one possible explanation for the separation of A and B on the DEAE-Sepharose column was that they were diastereomers. Racemisation of amino acids, especially histidine, is often associated with peptide synthesis (see figure 2.7), therefore the chiral integrity of the amino acids in A and B was checked by carrying out an enzymatic digestion. Three enzymes, pepsin, prolidase and aminopeptidase, all of which recognize only

L-amino acids, were used to hydrolyse the two synthetic materials into their constituent amino acids (see table 2.2).

Amino Acid	Expected Value	Bovine Ubiquitin	Synthetic A	Synthetic B
Asp	5	4.21	4.90	4.63
Thr	7	9.30	8.85	8.56
Ser	3	4.67	4.63	4.48
Glu	6	7.71	6.92	6.71
Pro	3	2.32	2.40	2.64
Gly	6	6.24	7.05	6.56
Ala	2	1.81	2.11	2.07
Val	4	4.27	4.17	4.00
Met	1	Met O	Met O	Met O
Ile	7	6.87	6.65	6.40
Leu	9	10.35	10.66	10.62
Tyr	1	0.98	1.46	0.37
Phe	2	2.70	2.50	2.49
His	1	1.05	0.92	1.04
Lys	7	6.36	6.30	6.00
Arg	4	3.95	3.56	3.50

Table 2.2 Total Enzymatic Digestion of A, B and Bovine Ubiquitin

Although amino acid analysis of these enzymatic hydrolyses did not in general give as accurate a set of ratios as analysis of the corresponding acid hydrolyses, it did however provide some valuable information. For example, the histidine content in both A and B was clearly as required, indicating no racemisation of this amino acid had in fact taken place. Evidently the precautions taken to minimize this had been successful. Furthermore, asparagine and glutamine residues were not hydrolysed to their constituent acids, therefore only the true Asp and Glu contents were observed. This made such an enzyme hydrolysis complementary to the acid hydrolysis.

Enzymatic hydrolysis is not only sensitive to the chirality of amino acids, but also to their isomorphism. For example, during cleavage aspartic acid residues can,

under extreme conditions, rearrange to β-aspartamate (**50**) via the succinimide intermediate (**49**) shown in figure 2.15 [194].

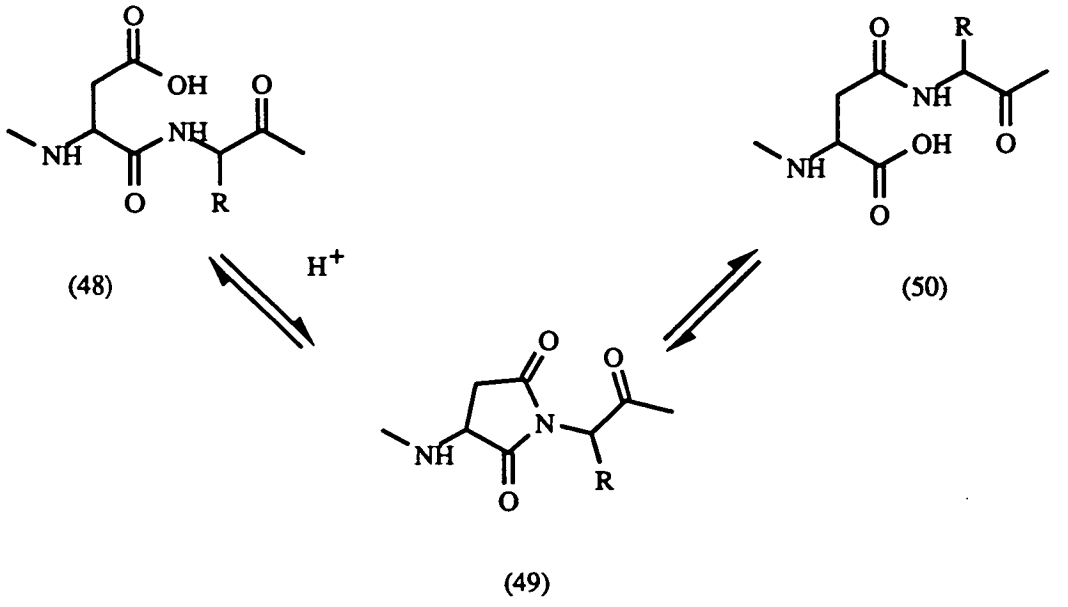


Figure 2.15 β-Aspartyl Formation

Such a rearrangement cannot be detected by acid hydrolysis of the peptide since aspartic acid is still the hydrolysis product. However, an enzyme at neutral pH would not recognize the β-amino acid and so could not fully digest such a peptide. Neither synthetic A nor synthetic B is conspicuously low in aspartic acid residues relative to bovine ubiquitin, suggesting this type of rearrangement had not occurred in either case (see table 2.2). This contention was given further credibility by the remarkable similarity of the 2D NMR spectra of synthetic A and B. Figure 2.16 shows the amide to CH crosspeak region in the TOCSY spectrum of A superimposed upon the same region in the TOCSY spectrum of B. This part of the spectrum, known as the fingerprint region, is extremely sensitive to tertiary structure and such a high level of coincidence is strong evidence indeed for both materials having an identical conformation under these NMR conditions.

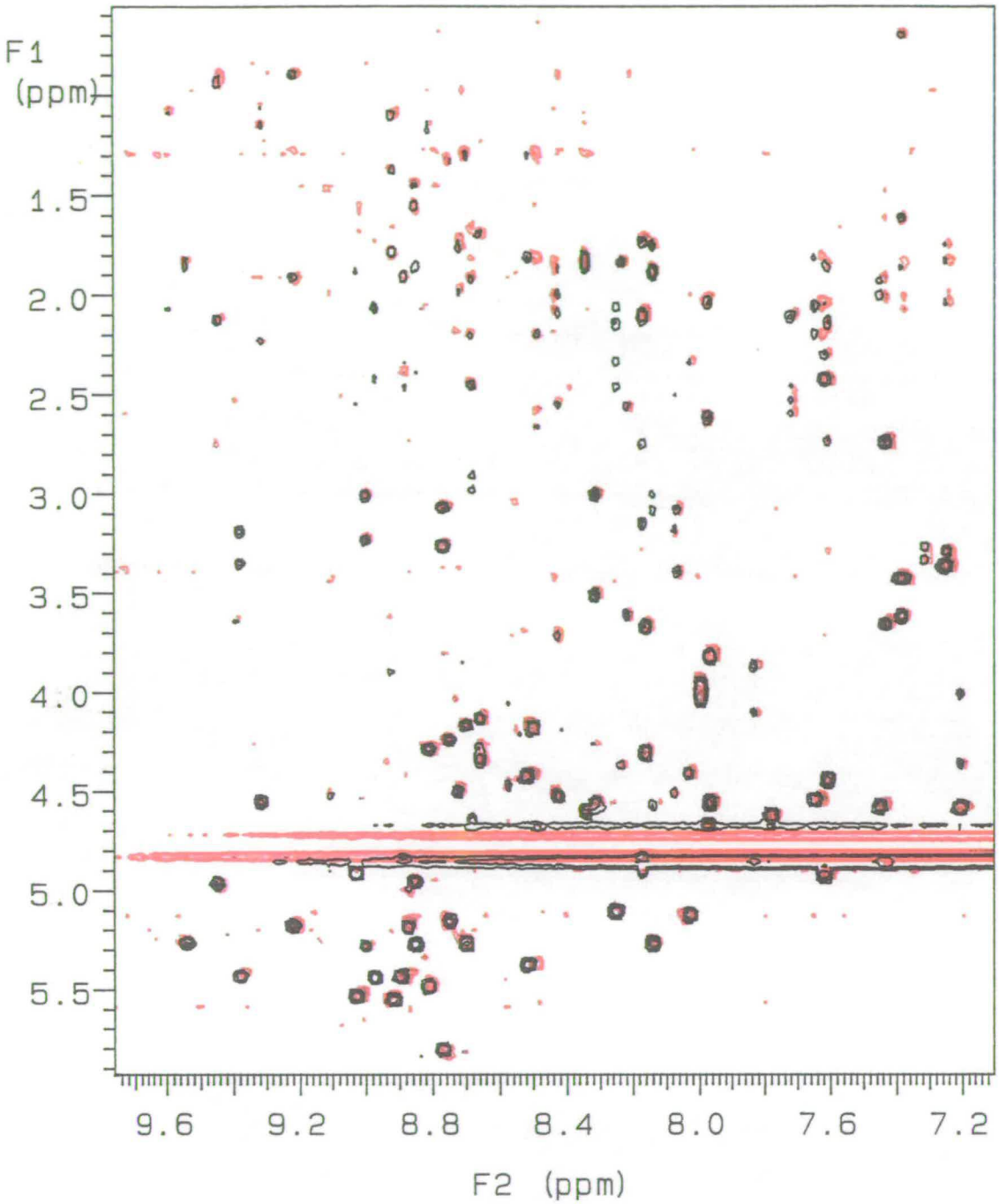


Figure 2.16 **Comparison of the TOCSY Spectrum of A (red) and B (black). Spectra obtained in 90% H_2O: 10% D_2O at pH 4.8, 50°C**

Having as far as possible demonstrated that synthetic A and B were chemically identical, the observation that at pH 4.8 they were also conformationally equivalent presented quite a dilemma. It seemed likely that the answer to this problem lay in the conditions used in the final anion exchange step of the purification. The high

pH of 9.3 may well have been enough to partially denature the synthetic material and hence account for the two peaks observed. If this were the case then not only would both materials be chemically identical, but at reduced pH the partially denatured state may well refold. This could of course explain why A and B had the same TOCSY spectrum at pH 4.8.

To investigate this possibility a sample of synthetic A was dissolved in a buffer containing 20 mM piperazine at pH 9.6 and a small sample of this solution periodically analysed by anion exchange chromatography. This involved loading a fixed volume of this solution onto an analytical Mono Q FPLC (Fast Protein Liquid Chromatography) column pre-equilibrated with starting buffer, thereafter applying a salt gradient of 0 to 1 M NaCl to elute the protein. Our hypothesis was vindicated by the gradual appearance of a second peak on each successive FPLC trace (see figure 2.17(a)).

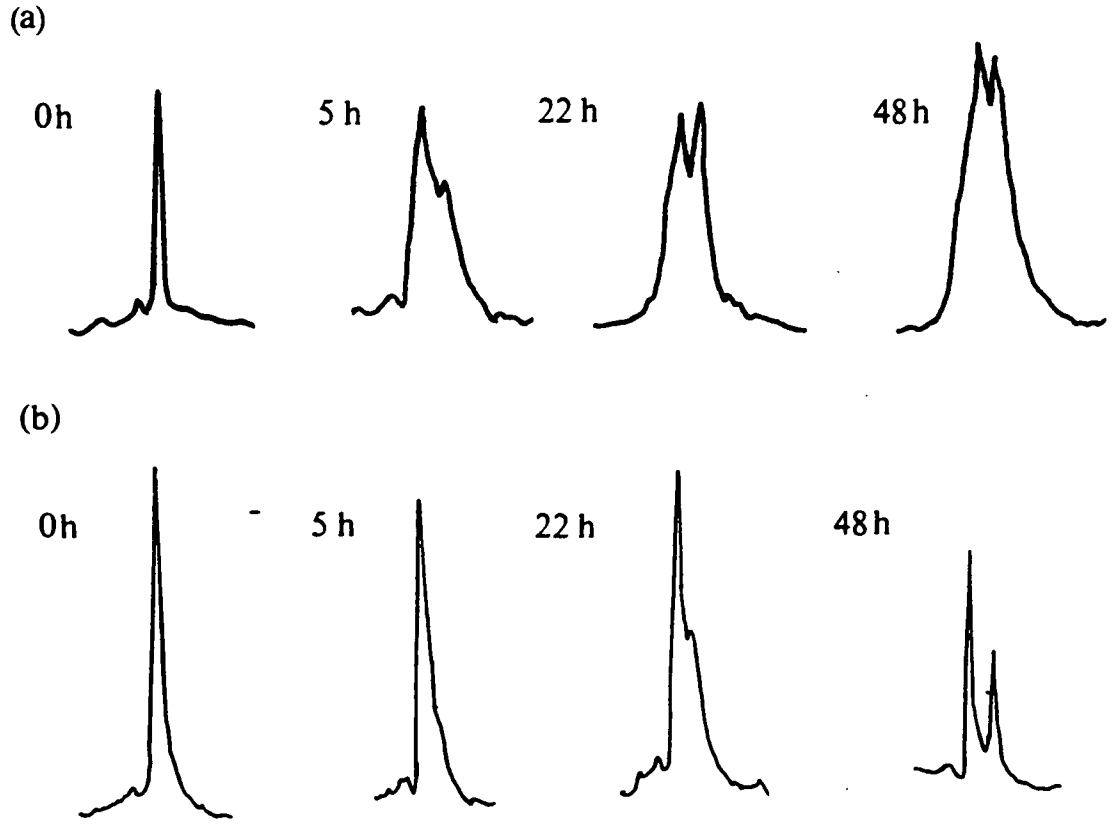
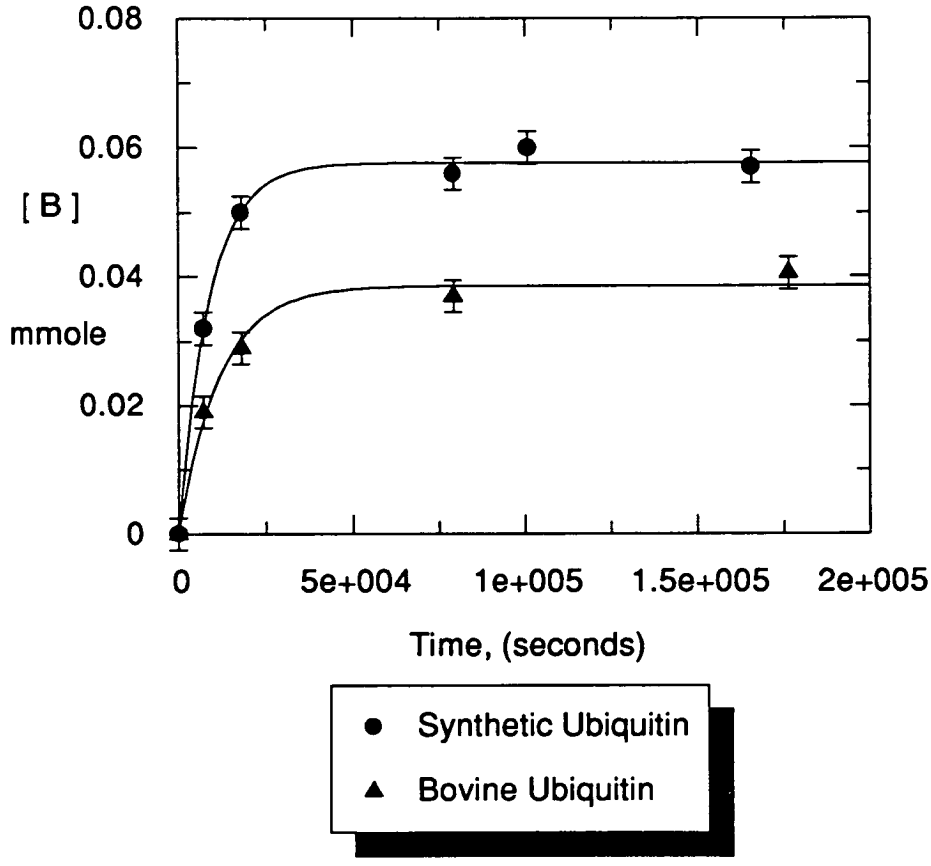


Figure 2.17 FPLC Monitored Equilibration of Synthetic (a) and Bovine Ubiquitin (b) to the A and B Forms

By estimating the relative areas of the two peaks and knowing the total number of moles injected onto the column, it was possible to show that this equilibrium process initially obeyed first order kinetics (see graph 2.1 and table 2.3). Moreover, the second peak completely disappeared after dialysing the protein solution back to pH 7.0, indicating both the reversible nature and pH dependence of the denaturation process. That synthetic material B was indistinguishable from material A under these FPLC conditions surely affirms that its initial isolation from the DEAE column was as a result of the high pH conditions used. Subsequent exposure to neutral conditions would then have allowed it to revert back to the A form.



Graph 2.1 [B] Form vs Time for Synthetic and Bovine Ubiquitin

Purified bovine ubiquitin was also shown to undergo an equilibration process when dissolved in 20 mM piperazine at pH 9.6 (see figure 2.16(b)). This appeared to be a first order process approaching equilibrium and gave a rate constant similar to that of synthetic ubiquitin (graph 2.1 & table 2.3). The position of the equilibrium

appears to be further in favour of the less stable form for synthetic ubiquitin (K = 0.97) than for bovine ubiquitin (K = 0.50) implying that the former is more easily denatured than the latter. While it is possible that synthetic ubiquitin has a somewhat less stable conformation than bovine ubiquitin, it should be stressed that this experiment was designed primarily to show an equilibrium did exist and not to accurately pin-point the position of such a process.

	$k\ s^{-1}$	K
Synthetic Ubiquitin	112 ± 0.73 e -6	0.97
Bovine Ubiquitin	85 ± 0.98 e -6	0.50

Table 2.3 Kinetic and Thermodynamic Data on the Equilibration of Bovine and Synthetic Ubiquitin

From this FPLC work it is fair to conclude that synthetic materials A and B are in fact constitutionally identical, albeit initially in different conformations. This means that the overall yield of synthetic ubiquitin was 93.5 mg or 4.4%, a ten fold increase in yield over any previous attempts at the construction of the protein.

Having resolved the A and B problem some additional characterization of the synthetic ubiquitin was performed. A preliminary mass determination using laser desorption mass spectrometry (LDMS) had given satisfactory results for synthetic ubiquitin, however it was felt that electrospray mass spectrometry could perhaps afford an even more accurate mass determination on the polypeptide (see figure 2.18). Unlike LDMS this technique does not rely on time-of-flight as a means of mass determination instead employing the more familiar quadrapole detection system. Although quadrapole detection becomes increasingly less useful as the sample gets larger, it will usually yield a more accurate result than time-of-flight detection for a protein the size of ubiquitin due to multiple charging. This technique did indeed prove to be superior to LDMS, providing a molecular mass of 8564.8 Da for synthetic ubiquitin compared to a theoretical mass of 8564.9 Da. Furthermore, apart from the sodium salt at 8586 Da there were no signs of any impurities being present in the sample.

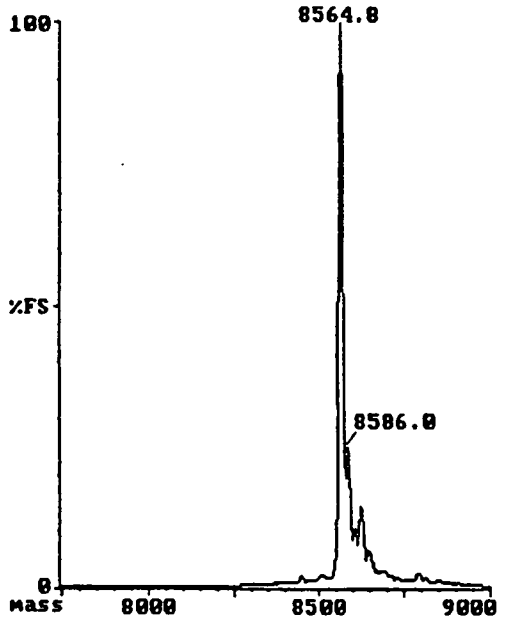


Figure 2.18 Electrospray Mass Spectrometry on Synthetic Ubiquitin Showing the Calculated Mass at 8564.8 Da plus the Sodium adduct at 8586 Da

Extensive characterization of the primary structure of the protein was carried out by automated Edman degradation which was fortunate in sequencing through to residue 67 without ambiguity. This technique alone, starting as it does at the portion of the molecule synthesised last, almost suffices in proving the primary sequence was free of deletions.

Peptide	Peptide sequence[a]
T-0	Ala-Lys (28-29)
T-1	Ile-Gln-Asp-Lys(30-33)
T-2	Leu-Arg-Gly-Gly (73-76)
T-3	Thr-Leu-Thr-Gly-Lys (7-11)
T-4	Gln-Leu-Glu-Asp-Gly-Arg (49-54)
T-5	Glu-Gly-Ile-Pro-Pro-Asp-Gln-Gln-Arg (34-42)
T-6	Thr-Leu-Ser-Asp-Tyr-Asn-Ile-Gln-Lys (55-63)
T-7	Leu-Ile-Phe-Ala-Gly-Lys (43-48)
T-8	Met-Gln-Ile-Phe-Val-Lys (1-6)
T-9	Glu-Ser-Thr-Leu-His-Leu-Val-Leu-Arg (64-72)
T-10	Thr-Ile-Thr-Leu-Glu-Val-Glu-Pro-Ser-Asp-Thr-Ile-Glu-Asn-Val-Lys (12-27)

Table 2.4 Identification of the Peptides arising from Tryptic Digest of Ubiquitin. [a] Determined by amino acid composition. Residue numbers are given in parenthesis

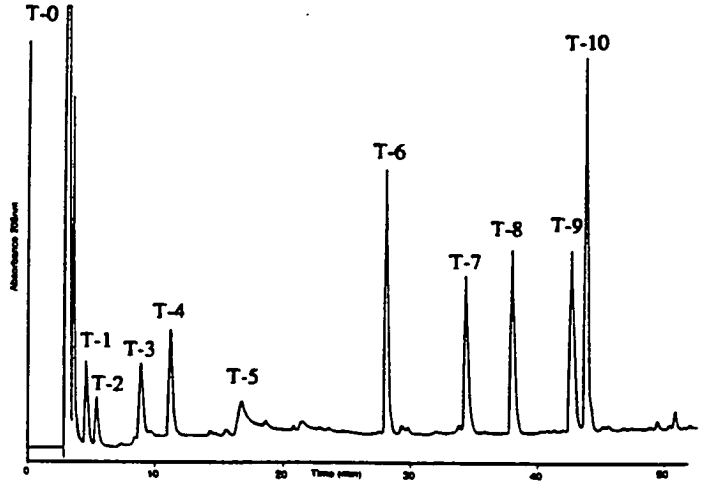


Figure 2.19 Tryptic Map of Synthetic Ubiquitin
Separation performed on a C_{18} Vydac HPLC Column

Finally, enzymic digestion of the synthetic material with trypsin [196] further demonstrated that its primary sequence matched that of ubiquitin. Isolation of the resulting peptide fragments by HPLC (figure 2.19), followed by amino acid analysis of these (table 2.4), confirmed the tryptic map was that of ubiquitin.

In conclusion, ubiquitin has been chemically synthesised and purified in 4.4% overall yield. Exhaustive chemical and biochemical characterization of this synthetic protein not only confirms that it is the required material, but also that it is >95% homogeneous.

2.4.3 Structural Analysis of Synthetic Ubiquitin

At the outset of this work it was known that simply constructing the ubiquitin primary sequence was not in itself enough, the synthetic protein would also have to adopt the same folded state as natural ubiquitin before the synthesis was deemed a success. Therefore the conformation of the synthetic ubiquitin was investigated using both classical techniques, as well as the more comprehensive X-ray crystallography and nuclear magnetic resonance approaches.

2.4.3(i) Ultra-Violet Absorption Spectroscopy

Ubiquitin contains a single tyrosine residue in a conformationally sensitive region of its tertiary structure, hence this can be used as a convenient probe into the conformation of the protein. In the crystal structure of natural ubiquitin the phenolic

group of Tyr-59 is known to hydrogen bond to the backbone nitrogen of Glu-51 which would of course raise the pKi of the phenolic group in the tyrosine. Ultra-violet absorption spectroscopy allows this pKi to be determined since the ionised phenolic group absorbs at a wavelength independent from the rest of the chromophores in the protein (see figure 2.20).

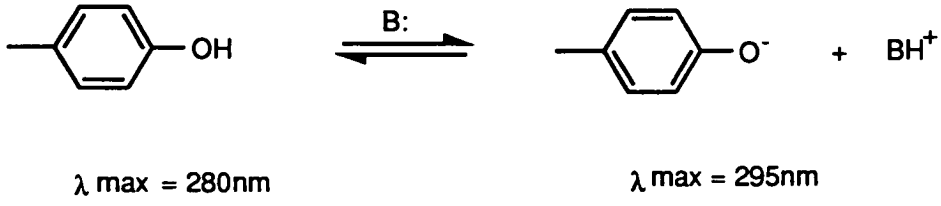
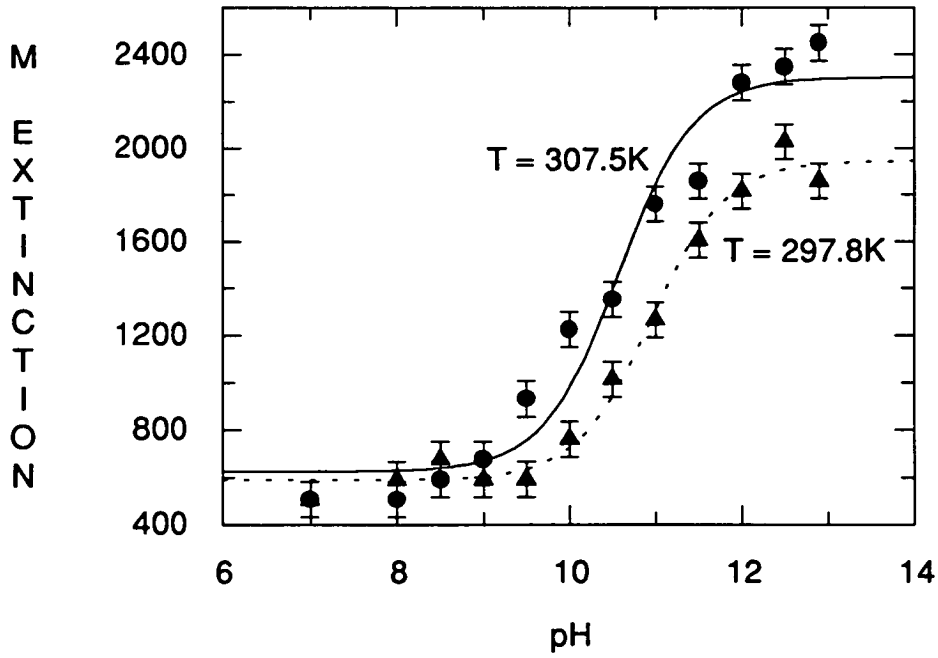


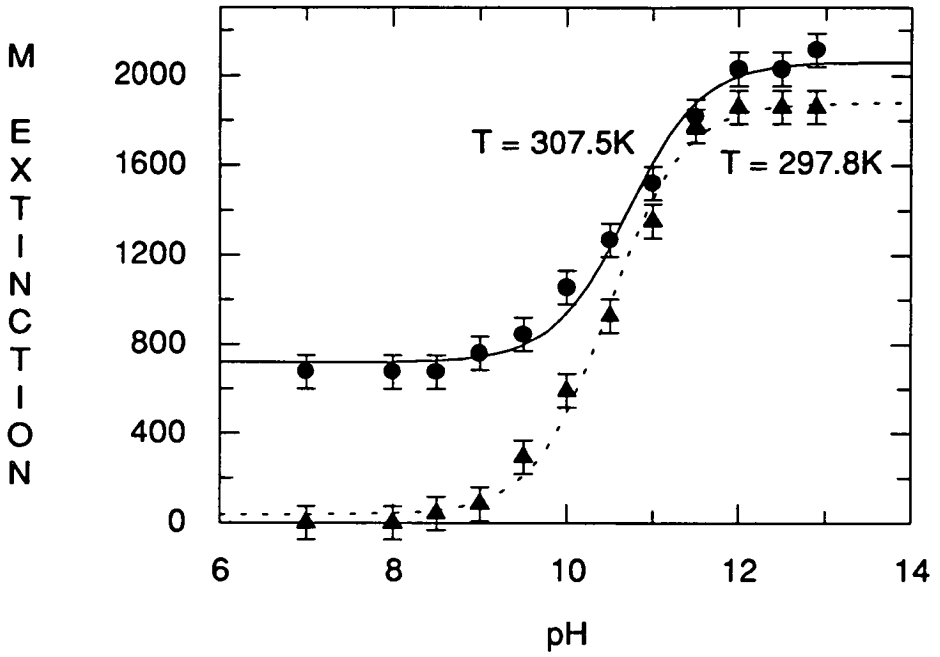
Figure 2.20 Ionisation of the Phenolic Group in Tyrosine

A simple base titration monitored by UV absorption at 295 nm affords the pKi of the tyrosine. At 298K the pKi of the tyrosine in purified bovine ubiquitin was found to be 11.00 (see graph 2.2) and this agrees with a value of 11.10 determined by Goldstein and co-workers [112].



Graph 2.2 Spectrophotometric Titration (295 nm) of the Phenolic Group in the Tyrosine of Purified Bovine Ubiquitin at Two Temperatures

At the same temperature the tyrosine in synthetic ubiquitin has a pKi of 10.50 (see graph 2.3). Although slightly lower than in bovine ubiquitin the tyrosine pKi in synthetic ubiquitin is still substantially higher than in several other proteins (see table 2.5) indicating the phenolic group is engaged to some degree in hydrogen bonding.



Graph 2.3 Spectrophotometric Titration (295 nm) of the Phenolic Group of Tyr-59 in Synthetic Ubiquitin at Two Temperatures

When these experiments are carried out at more than one temperature, the resulting pKi data can be used to calculate the heat of ionisation (ΔHi) of the phenolic group[197-199]. The heat of ionisation can be obtained from the change in pKi with temperature using relationship (1).

$$\Delta Hi = 2.303 * R * \left[\frac{\Delta pki}{\Delta (1/T)} \right] \qquad 1$$

A value of 10.6 kcal mol^{-1} was obtained for ΔHi of the tyrosine in synthetic ubiquitin compared with 10.8 kcal mol^{-1} for the tyrosine in bovine ubiquitin (see table 2.5).

Example	pKi, 298K	pKi, 308K	ΔHi, Kcal mol^{-1}
Synthetic Ub	10.50	10.25	10.6
Bovine Ub	11.00	10.75	10.8
BSA [197]	11.50	11.20	11.5
Insulin [198]	9.70	NA	7.5
Ribonuclease [199]	9.92	NA	7.0
Pepsin [197]	9.50	NA	6.0
Phenol [197]	9.78	NA	6.1
Tyrosine [197]	10.05	9.90	6.0
Tyr-Arg [199]	9.60	NA	6.0

NA = Not Available

Table 2.5 Thermodynamic Data on Synthetic and Bovine Ubiquitin

While this thermodynamic data suggests the tyrosine in synthetic ubiquitin is in a similar environment to the tyrosine in bovine ubiquitin, the slightly smaller value of ΔHi obtained for the former may be because the H-bond between the tyrosine and Glu-51 is not as strong as in the latter.

2.4.3(ii) Quenching Studies on the Tyrosine Fluorescence

Fluorescence spectroscopy can perhaps best exploit the solitary tyrosine probe in ubiquitin. It is a considerably more sensitive technique than UV spectroscopy and should be able to detect subtle changes in the immediate environment of the tyrosine fluorophore.

The fluorescence of a protein solution can be quenched by addition of an external heavy atom such as the iodide ion [200-201]. Collisions with these heavy atoms causes spin-orbital perturbation in the fluorophore. This in turn leads to intersystem crossing from an excited singlet state (from which fluorescence would normally occur) to a non-fluorescent triplet metastable state (see figure 2.21).

Intersystem crossing is usually a forbidden transition and for this reason will only occur when a heavy atom collides with the fluorophore [202]. Hence, the more collisions that take place, the more the fluorescence is quenched. Obviously fluorophores residing on the surface of a protein will have their fluorescence quenched to a greater extent than those buried and thus protected within core of a protein.

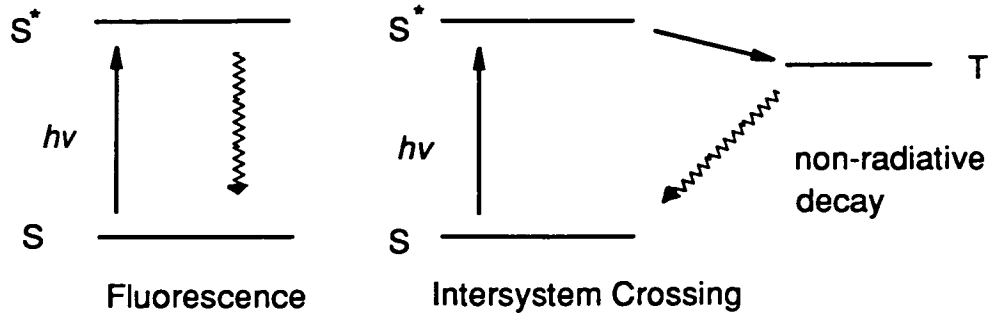


Figure 2.21 Principles of Fluorescence and Intersystem Crossing

If a protein contains a single fluorophore (eg ubiquitin) then the degree of quenching is related to the heavy atom concentration [HA] by the Stern-Volmer law (2).

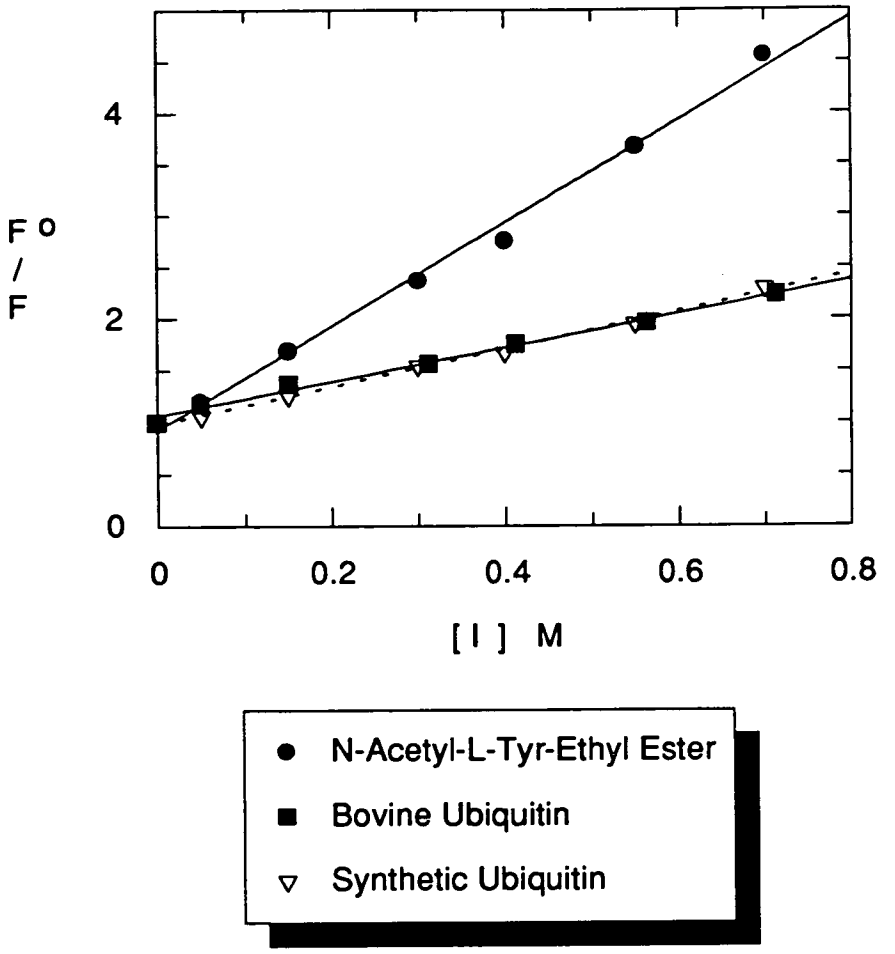
$$F^o/F = 1 + K_Q [HA] \qquad 2$$

Where F is the fluorescence at a particular heavy atom concentration, F^o is the fluorescence in the absence of any heavy atoms and K_Q is the Stern-Volmer constant. A plot of F^o/F *vs* [HA] should give a straight line of slope K_Q. Clearly the larger K_Q then the more exposed the fluorophore is to the bulk solvent.

The tyrosine fluorescence in both synthetic and purified bovine ubiquitin was quenched by gradually increasing the concentration of iodide ions in the protein solution. Graph 2.4 shows the resulting Stern-Volmer plots obtained for the two proteins, as well as for an N-acetyl-tyrosine ethyl ester (ATEE) control.

Example	K_Q M^{-1}
Synthetic Ub	1.82 ± 0.05
Bovine Ub	1.76 ± 0.08
ATEE	5.00 ± 0.167

Table 2.6 Experimentally Determined Stern-Volmer Constants (K_Q)



Graph 2.4 Stern-Volmer Plots

The experimentally obtained values of K_Q are shown in table 2.6. Synthetic and bovine ubiquitin have almost identical Stern-Volmer constants, indicating the tyrosine in each has a similar exposure to the bulk solvent. In addition, compared to ATEE, which corresponds to complete exposure to the solvent, the values of K_Q for the ubiquitin is quite small. This suggests the tyrosine in ubiquitin is at least partially buried in the structure. Indeed, this is observed to be the case in the crystal structure of the protein.

2.4.3(iii) Denaturation Studies on Synthetic and Bovine Ubiquitin

The sensitivity of fluorescence spectroscopy to changes in protein structure makes it an ideal technique for monitoring the behaviour of a protein exposed to varying concentrations of a denaturant, such as guanidinium hydrochloride (GdmCl). A protein in its native conformation will generally fluoresce with different intensity to

that same protein in a denatured state. This is because in the folded state the fluorescence of a protein is partially quenched by amide bonds in the surrounding structure. In the unfolded state such bonds are often no longer in the proximity of the fluorophore and hence the fluorescence is no longer quenched. Note that in some cases (eg, barnase [142]) this is the other way round leading to a lower fluorescence intensity in the denatured state than in the native form.

Gradually increasing the [GdmCl] gives rise to a characteristic fluorescence denaturation curve from which it is possible to abstract useful thermodynamic information on the protein. At any particular denaturant concentration the equilibration constant for unfolding, Ku, can be calculated from equation (3),

$$Ku = (F_N - F) / (F - F_U) \qquad 3$$

where F is the observed fluorescence and F_N and F_U are the values of the fluorescence for the native and unfolded form of the protein. The free energy of unfolding, ΔGu, is related to the concentration of the denaturant by equation (4)[203].

$$\Delta Gu = \Delta G_{H2O} - m\,[\text{denaturant}] \qquad 4$$

Given that ΔGu can also be calculated from the fundamental relationship (5),

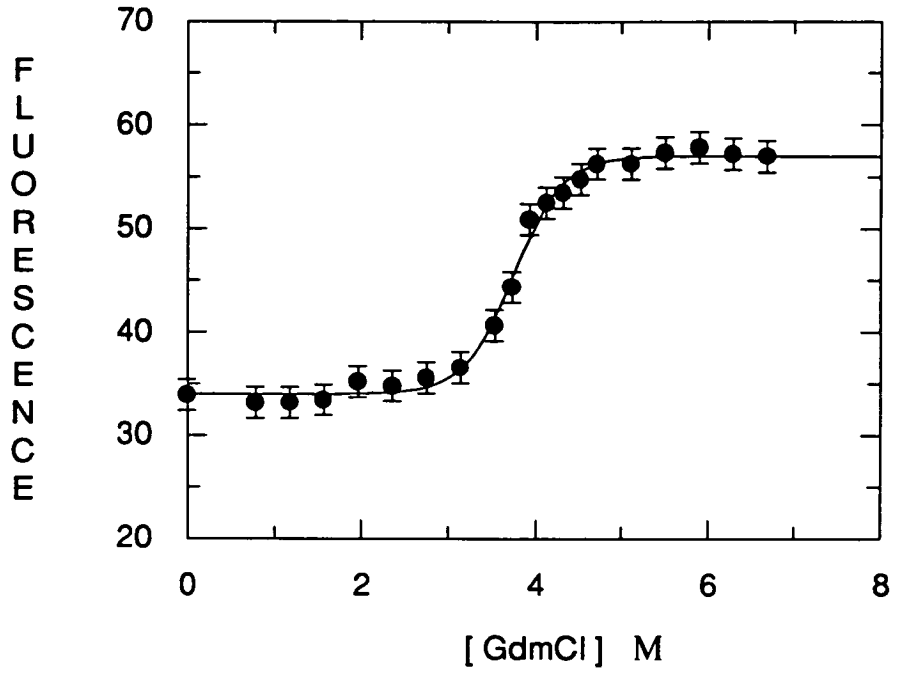
$$\Delta Gu = -RT\ln K_U \qquad 5$$

then a plot of ΔGu *vs* [denaturant] will afford the free energy of stabilization, ΔG_{H2O}, if extrapolated back to the y axis. The free energy of stabilization can also be obtained by non-linear regression analysis (Grafit, Microsoft) using equation (6) which is derived from equations (3-5).

$$F = FN - [FN - FU] \Big/ 1 + EXP\left[\frac{m^* G - [D]}{8.314 * T}\right] \qquad 6$$

Graph 2.5 shows the change in tyrosine fluorescence (measured at 303 nm) produced upon addition of GdmCl to a synthetic ubiquitin solution. The nature of

this denaturation curve reveals that the tyrosine fluorescence in synthetic ubiquitin is partially quenched in the folded state.



Graph 2.5 GdmCl Induced Unfolding of Synthetic Ubiquitin

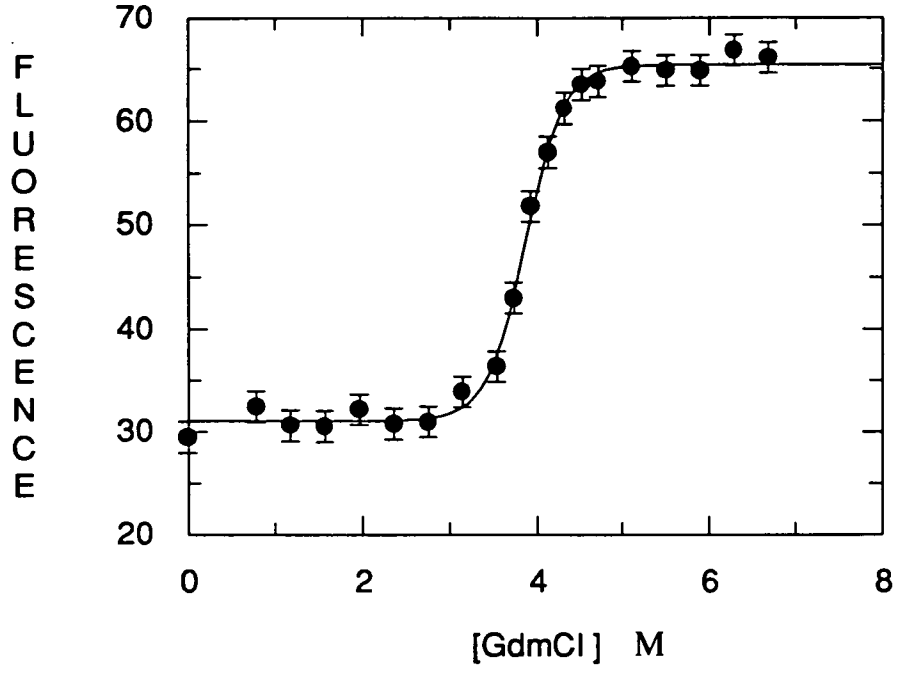
Similarly, graph 2.6 shows the fluorescence unfolding curve obtained for purified bovine ubiquitin under identical conditions. Clearly the tyrosine fluorescence is also quenched in the folded state of the natural protein. Non-linear least squares analysis of these data sets (using equation 6) yielded the free energy of stabilization, ΔG_{H2O}, for the synthetic and bovine materials (see table 2.7).

Example	ΔG_{H2O} kcal mol^{-1}
Synthetic Ubiquitin	-8.1 ± 0.8
Bovine Ubiquitin	-8.0 ± 1.0
Barnase [142]	-9.4

Table 2.7 Free Energy of Stabilization (ΔG_{H2O}) Data

The value of ΔG_{H2O} obtained for synthetic ubiquitin compares favourably with that of its natural counterpart, again pointing to the two materials having a similar conformation in solution. Furthermore, the free energy of stabilization for ubiquitin

is not much different from that calculated for barnase [142]. This is especially encouraging given that barnase, like ubiquitin, is a relatively small protein with a pronounced hydrophobic core and no structure stabilizing disulphide bridges.



Graph 2.6 GdmCl Induced Unfolding of Purified Bovine Ubiquitin

2.4.3(iv) Circular Dichroism Studies on Synthetic and Bovine Ubiquitin

Circular dichroism (CD) spectroscopy is a technique which unlike UV absorption or fluorescence spectroscopies, provides quantitative information on the secondary structure content of a protein. The fundamental feature peculiar to CD is that the protein solution is exposed to plane polarised light, which can be regarded as a mixture of right- and left-handed circularly polarised light. If the protein contains structural elements in which the individual chiral chromophores are arranged non-randomly (eg, a helix), then at any given wavelength these left- and right-handed components will be absorbed to different extents. In effect, this results in the light emerging from the protein being tilted through an angle θ relative to the polarisation of the incident light. In general the more secondary structure a protein contains then the larger this angle will be.

As in the previous structural studies, we directly compared synthetic ubiquitin with purified bovine ubiquitin. In work carried out at the Scottish circular dichroism

facility in Stirling University [204], the far UV CD spectra of synthetic and bovine ubiquitin were found to be remarkably similar (see figure 2.22).

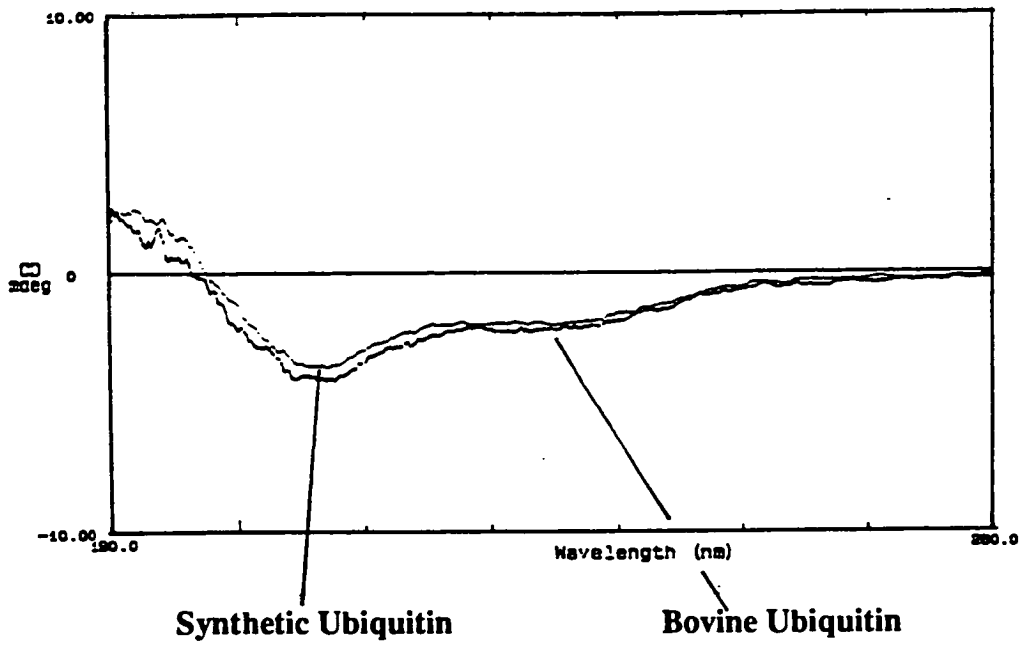


Figure 2.22 Far UV CD Spectra of Synthetic and Bovine Ubiquitin in 50 mM Boric Acid at pH 7.5

The secondary structure content of synthetic and natural ubiquitin were determined using the CONTIN procedure devised by Provencher and Glockner [205]. This involves comparing the 190-240 nm region in the CD spectrum of the sample with a linear combination of the CD spectra of 16 proteins whose secondary structures are known from X-ray crystallography. In general, this technique provides more reliable results than direct comparison with a single reference spectrum [206]. Table 2.8 summarises the secondary structure contents calculated using this procedure.

Example	α-helix	β-sheet	Remainder
Crystal Structure	14.4%	42%	43.5%
Bovine Ub, pH7.5	15 ± 2.5%	57 ± 2.6%	28 ± 4.5%
Synthetic Ub, pH7.5	16 ± 1.1%	53 ± 1.9%	32 ± 2.1%
Bovine Ub, pH9.3	17 ± 0.7%	32 ± 1.3%	51 ±1.2%
Synthetic Ub, pH9.3	20 ± 0.9%	32 ± 1.8%	48 ± 1.6%

Table 2.8 Secondary Structure Contents Derived From CD

From these results synthetic and bovine ubiquitin would appear to have very similar secondary structure compositions at pH 7.5. In each case the calculated proportion of α-helix in the structure is much closer to the known value (from the crystal structure) than is the rather high β-sheet content.

Having established that CONTIN treatment of the CD spectra gives reasonably accurate secondary structure data, we decided to investigate the effect of high pH on the CD spectra of synthetic and bovine ubiquitin (see figure 2.23).

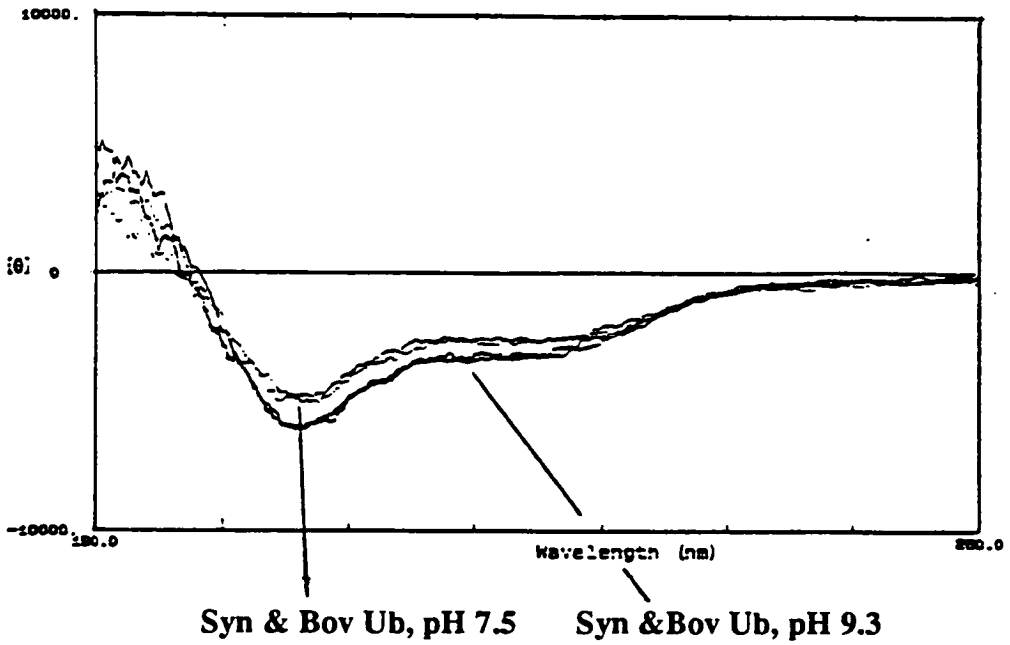
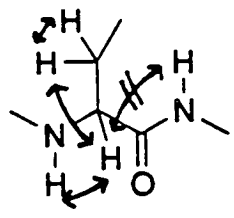


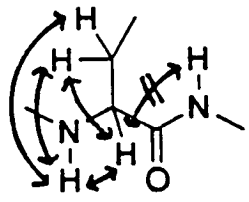
Figure 2.23 Effect of pH on the Far UV CD Spectra of Synthetic and Bovine Ubiquitin

Previous studies had revealed a partial denaturation of ubiquitin at high pH (section 2.4.2). This denaturation apparently manifests itself as a slight increase in α-helicity and a substantial drop in β-sheet content when the CONTIN procedure is applied to the CD spectra of synthetic and natural ubiquitin (table 2.8). Although the small rise in the α-helix content is a rather curious observation, the considerable loss of β-sheet at the high pH was consistent with the formation of a denatured state. Indeed, a similar drop in β-sheet content was associated with the alcohol induced molten globule state of ubiquitin reported by Williams *et al*[134].

2.4.3(v) NMR Spectroscopy Studies on Synthetic Ubiquitin

By far the most comprehensive way of studying the solution conformation of a peptide or protein is to use NMR spectroscopy. Present day high-field NMR spectrometers can yield considerable information on proteins up to around 15 kDa in size. This is in part due to the resolution of these instruments and also because of the ascendancy of multi-pulse experiments which afford invaluable information on the interconnectivities of individual resonances in a spectrum. Such two dimensional NMR experiments not only allow complete assignment of complex spectra like those of proteins, but also facilitate the identification of secondary and tertiary structure within these molecules. Three of the most commonly used 2D NMR experiments are illustrated along with the information they provide in figure 2.24.





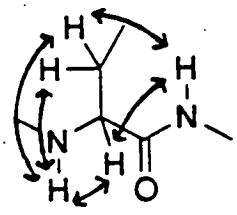


Figure 2.24 Three commonly used 2D NMR Experiments and the Connectivity Information they yield

All 2D NMR experiments have a variable delay time (t_1) between the first two pulses. Gradually increasing the length of t_1 in standard increments produces a corresponding set of free induction decays (FIDs) which when Fourier transformed in two dimensions give rise to the familiar 2D NMR spectrum. A comprehensible treatment of the theory behind 2D NMR is given in the book by Derome [207].

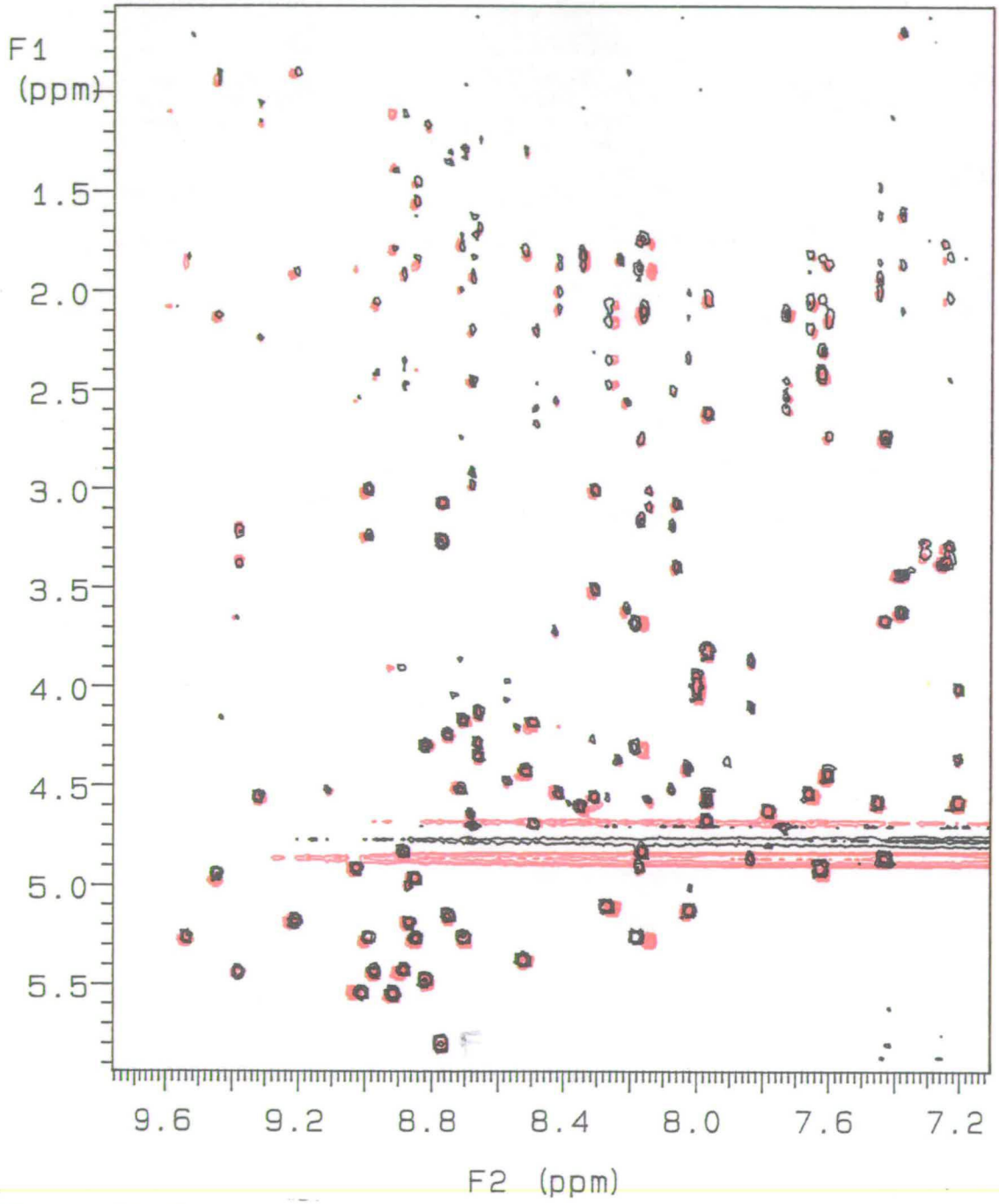


Figure 2.25 **Region From TOCSY Spectra of Synthetic (Red) and Bovine Ubiquitin (Black) obtained in 90% H_2O:10% D_2O, pH4.8, 50°C**

There have been two thorough 2D ^1H NMR studies carried out on natural ubiquitin[4,5] and together these greatly eased the proton NMR assignment of the synthetic material. As in all previous studies, synthetic ubiquitin was directly compared with purified bovine ubiquitin, thereby expediting analysis of its solution conformation.

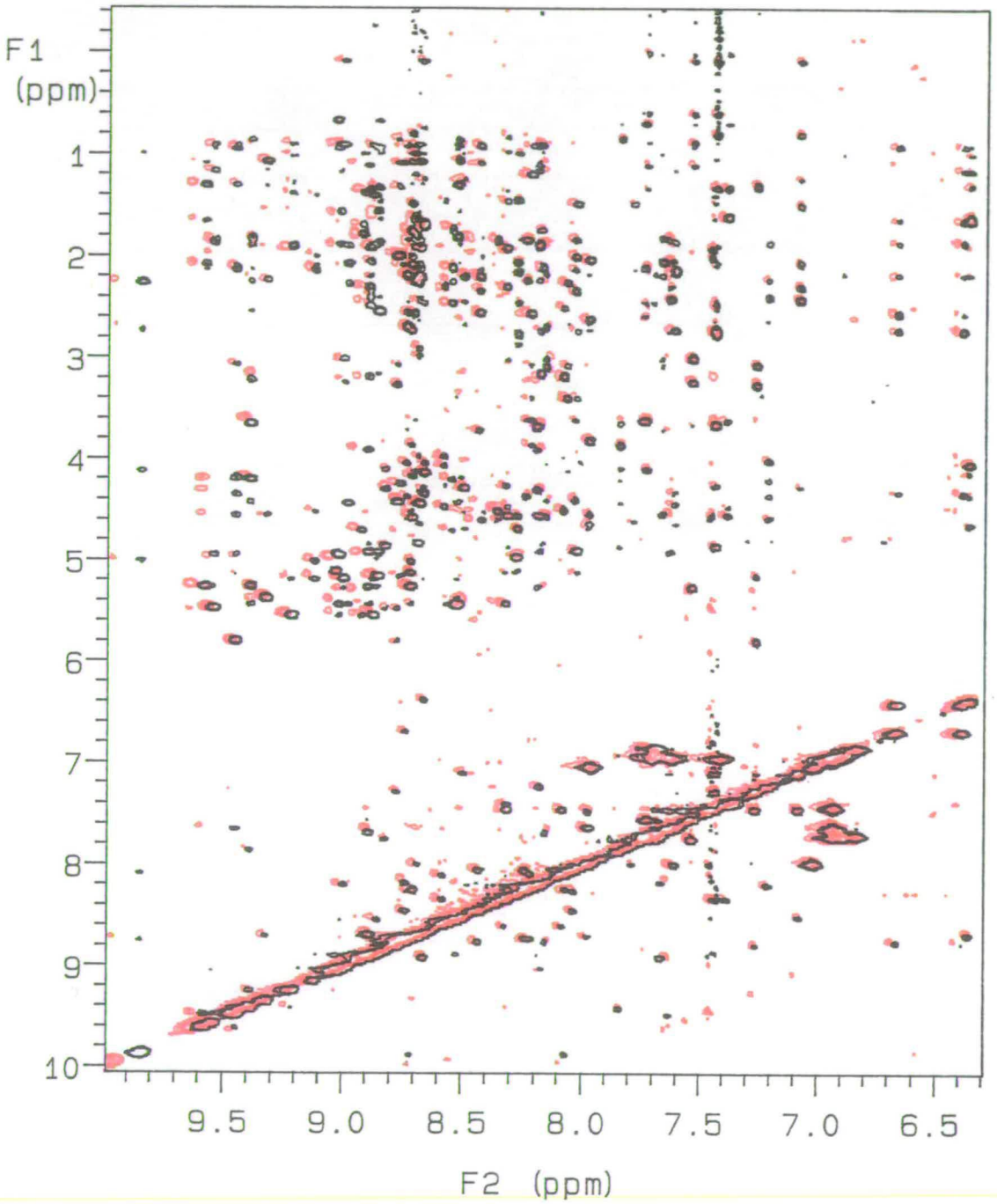


Figure 2.26 **Region From the NOESY Spectra of Synthetic (Red) and Bovine Ubiquitin (Black), 90/10% H_2O/D_2O at pH4.8, 50°C**

A TOCSY experiment was performed on both synthetic and bovine ubiquitin using conditions identical to those in the original papers [4,5]. Figure 2.25 shows a superimposition of the αNH - CH crosspeak regions in the two spectra. This part of a TOCSY spectrum is very sensitive to the solution conformation of a protein and for this reason is often known as the fingerprint region [121]. Consequently, the exceptional level of coincidence between these two spectra was strong evidence that synthetic ubiquitin had adopted the native conformation. Furthermore, a direct comparison with the natural ubiquitin ^1H assignment [4,5] allowed 529 out of the 579 observable protons in synthetic ubiquitin to be easily identified (this data is tabulated in appendix 1).

An even more precise measure of a protein's solution conformation is its NOESY spectrum. This experiment, which affords all important through space interactions, was performed on both synthetic and natural ubiquitin using conditions equivalent to those employed in previous studies [4,5] (see figure 2.26). Again, the spectrum obtained for synthetic ubiquitin was more or less indistinguishable from that acquired for purified bovine ubiquitin. Given that the presence of a crosspeak in a NOESY spectrum is very much dependant upon the structure of the molecule, such a high degree of congruence proved conclusively that both materials had an identical tertiary structure in solution. Upon mixing natural and synthetic ubiquitin only a single set of NMR crosspeaks were obtained in a TOCSY experiment, this is effectively the ultimate piece of evidence for the equivalence of the two materials.

2.4.3(vi) X-Ray Crystallography Studies on Synthetic Ubiquitin

X-ray crystallography, like NMR spectroscopy, can provide absolute information with regard to the conformation of a protein. Unlike NMR, this technique is not limited to the study of relatively small polypeptides, instead being applicable to proteins orders of magnitude larger. However, the protein must be available in a crystalline form, a precondition that is often as much an obstacle as size is to NMR spectroscopy. Note, there is no rule that the structure in the crystal matrix should match that of the protein in solution. Ubiquitin is a notoriously difficult protein to crystallize, to date this has only been achieved by the group responsible for the original crystal structure [114]. It was therefore not surprising that efforts to obtain crystals of synthetic ubiquitin were, in the first instance, unsuccessful. Fortunately, we were able to grow good crystals of the synthetic protein by seeding a synthetic ubiquitin solution with a single crystal of natural ubiquitin (kindly supplied by Dr

W. J. Cook of the University of Alabama Medical School). Synthetic ubiquitin crystals which grew independently of the seed crystal were then used as seeds in further crystallization experiments. This effectively minimized the risk of contamination with the natural ubiquitin crystal, as well as yielding larger synthetic crystals typically 0.3 mm in diameter. An example of the type of orthorhombic synthetic ubiquitin crystals obtained is shown in figure 2.27.

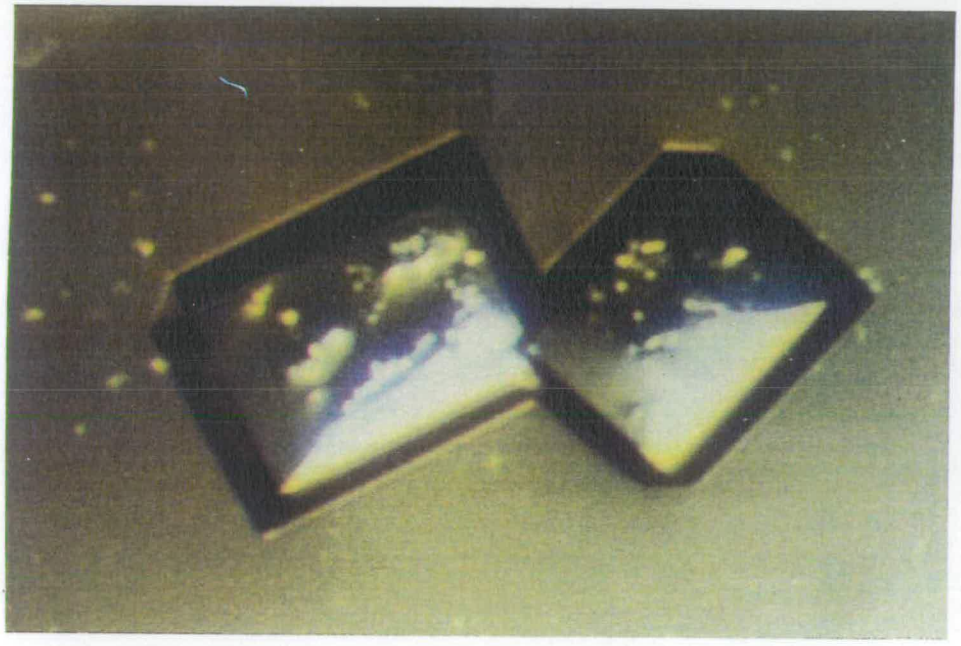


Figure 2.27 Crystals of Synthetic Ubiquitin

Diffraction of these crystals was carried out by Sawyer and his team in the Biochemistry Department at the University of Edinburgh [208] (see appendix 2 for details). In order to facilitate a more convenient comparison with synthetic ubiquitin, diffraction data from one of the natural crystals was also collected. Initial studies revealed that synthetic ubiquitin had crystallized isomorphously in the same space group ($P2_1\ 2_1\ 2_1$) as the natural protein and with almost identical unit cell parameters, a=50.83A°, b=42.74A°, c=28.98A° (natural ubiquitin a=50.73A°, b=42.74A°, c=28.78A°).

A detailed comparison of the synthetic and natural ubiquitin structure factors with those calculated from the published model of natural ubiquitin (1UBQ in

Brookhaven database) revealed the data set of the synthetic material (R factor=18.3%) conformed to the literature model slightly better than did the data set acquired for the natural crystals (R factor=19.8%). Nevertheless, direct comparison of the synthetic and natural experimental data sets confirmed the high overall similarity of the structures ($R_{SYN-NAT}$ =10.4%).

Refinement of the synthetic electron density map was subsequently performed using the molecular model of native ubiquitin in a molecular replacement procedure. Least-squares refinement was carried out and the results inspected and adjusted using the molecular graphics program FRODO [209]. After 6 rounds of refinement the synthetic structure had converged to R=16.5% or 1.7A^{o} resolution. Note that the published crystal structure for natural ubiquitin is at 1.8A^{o} resolution. The hydrogen bond network and dihedral angles of the main chain were very similar to the native ubiquitin starting point. Figure 2.28 shows the peptide backbone of synthetic ubiquitin superimposed upon that of its natural counterpart.

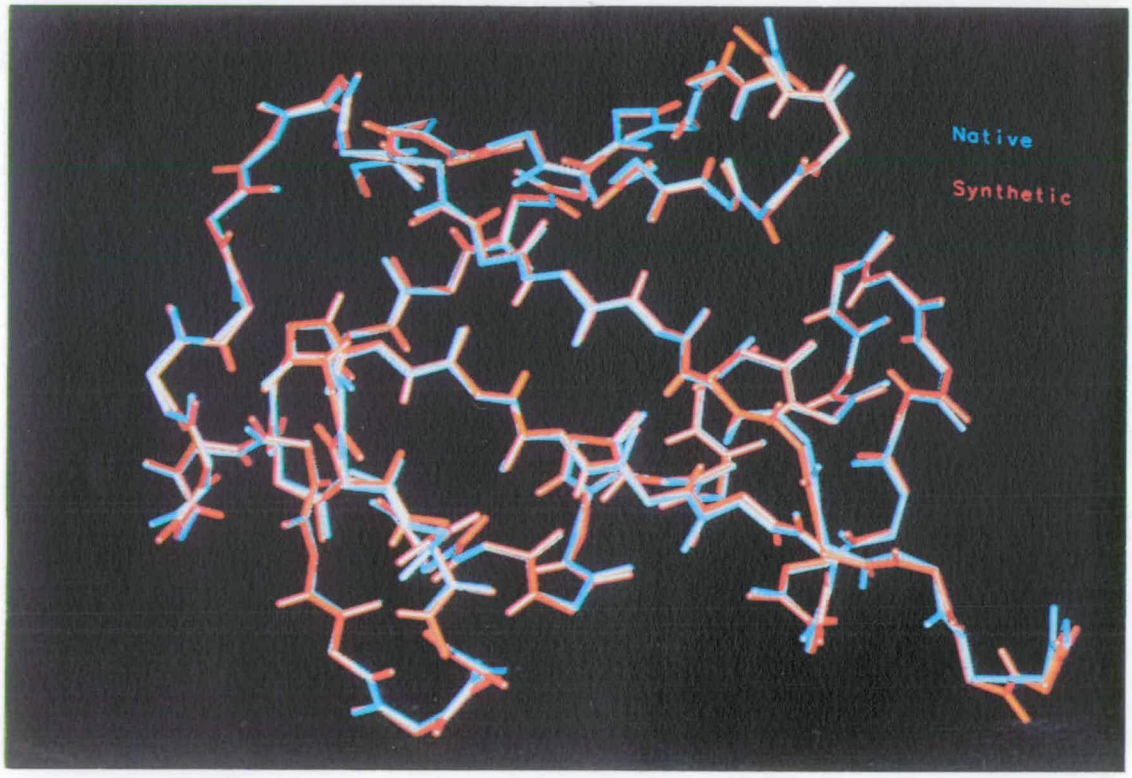


Figure 2.28 Comparison of the Crystal Structures of Synthetic and Natural Ubiquitin

The bulk of the side-chain positions were concordant with those in the literature crystal structure (see figure 2.29). However, the side-chains of five branched residues (Leu-15, Ile-44, Arg-54, Leu-69 & Arg-72) were slightly rotated relative to the literature model. Each of these side-chains is flexible and it is entirely reasonable that under the experimental conditions they can adopt essentially isoenergetic conformations.

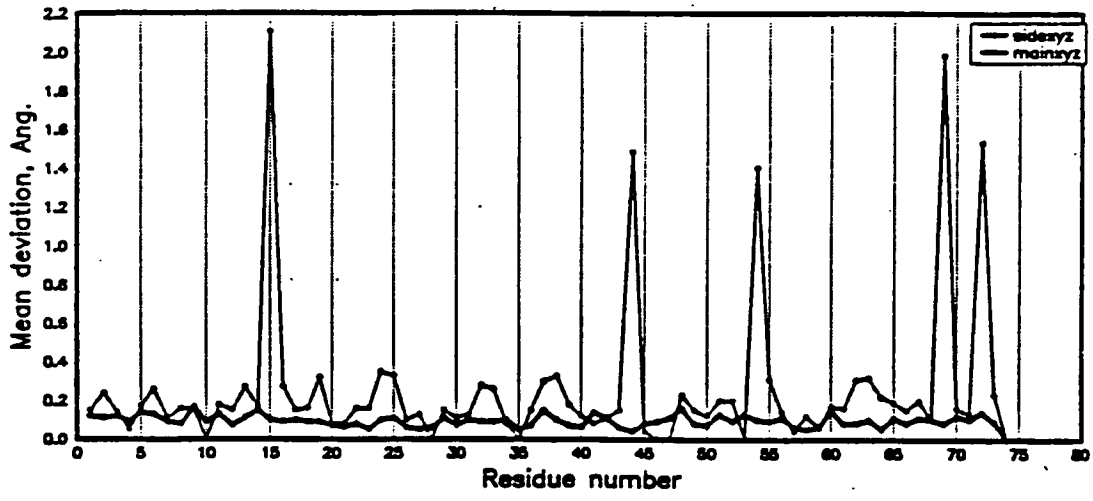


Figure 2.29 Mean Deviations of the Side- and Main-Chain Positions in Synthetic Ubiquitin from the Natural Model

In one further departure from the published structure, 82 water molecules were located in the course of the refinement, 50 of which were from the original 58 in the starting model. Most of the additional 32 water molecules were added to the bulk solvent in between the neighbouring protein molecules in the unit cell.

2.4.4 Protein Conjugation Activity of Synthetic Ubiquitin

The *in vitro* protein conjugation activity of synthetic ubiquitin was assayed [210]. Human placental tissue was initially lysed in phosphate buffer and the supernatent fractionated on a DE-52 Whatman anion exchange column at pH 7.2. The resulting high salt fraction II (FII), containing the components of the ubiquitin ligase system but importantly not ubiquitin, then served as the basis of the assay. Synthetic ubiquitin, radiolabelled with ^{125}I, was assayed for its ability to conjugate both endogenous and exogenous proteins.

Radiolabelled synthetic and bovine ubiquitin were incubated with human placental fraction II, supplemented with ATP, for 2 hours at 37°C. The contents of the assay were then separated by SDS-PAGE and the resulting ubiquitin-protein conjugates were visualized by autoradiography (see figure 2.30).

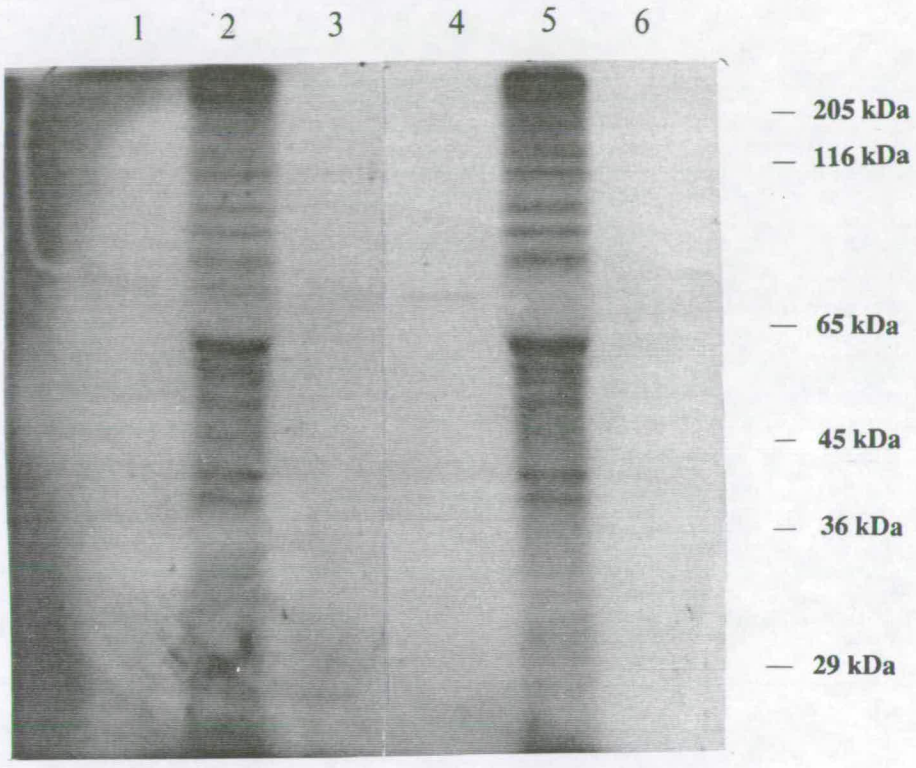


Figure 2.30 Endogenous Protein Conjugation Activities of Synthetic and Bovine Ubiquitin. Results from Autoradiography

lane1=Bov Ub **lane4=Syn Ub**

lane2=Bov Ub + FII + ATP **lane5=Syn Ub + FII + ATP**

lane3=Bov Ub + FII **lane6=Syn Ub + FII**

The two important lanes (2&5) contain a great number of high molecular weight conjugates, indicating both purified bovine ubiquitin and synthetic ubiquitin possess E1, E2 and E3 recognition activities. Although it is difficult to precisely determine their relative activities, it was found that from the same incorporated counts per minute (cpm) of ubiquitin, similar conjugation activities were obtained. In addition, the ATP dependence of this ubiquitin-protein conjugation activity was verified by starving the assay of ATP (lanes 3&6) resulting in no conjugates being formed.

In a further study, the ability of synthetic ubiquitin to conjugate exogenous lysozyme was investigated. In this assay, synthetic and purified bovine ubiquitin were incubated with human placenta fraction II, supplemented with ATP and lysozyme, for 60 minutes at 37°C (see figure 2.31)

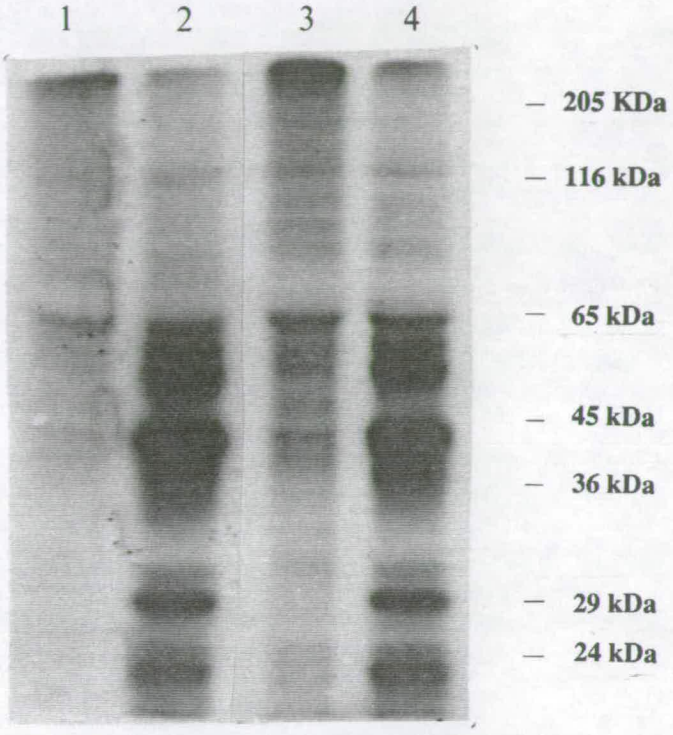


Figure 2.31 Exogenous Protein Conjugation Activities of Synthetic and Bovine Ubiquitin. Results from Autoradiography
lane1=Bov Ub + FII + ATP
lane2=Bov Ub + FII + ATP + 50μM lysozyme
lane3=Syn Ub + FII + ATP
lane4=Syn Ub + FII + ATP + 50μM lysozyme

Lanes 2&4 contain a series of bands corresponding to mono-, di-, tri-, etc ubiquitinated lysozyme, indicating both materials are able to conjugate the protein. Again synthetic and natural ubiquitin were apparently indistinguishable with regard to this activity.

Section 2.5 Synthesis and Studies on Ubiquitin ^{15}N-Gly^{5,26,30,43,50,56,67}, Ala ^{26,30} (Ubdes-Core)

2.5.1 Considerations Leading to the Design of Ubdes-Core

It has long been assumed, if not fully understood, that the hydrophobic core of a globular protein plays a large part in determining its conformation and mechanical stability. Recent studies by Lim and Sauer have revealed a number of components to the stability of a hydrophobic core, each contributing to the overall effect [146,211]. The most important feature is the hydrophobicity itself, ie the van der Waals interactions between neighbours. There are other fine tuning components which determine the precise structure and stability of the protein, namely; steric and volume constraints.

Allied to the hydrophobic effect in governing the folding of a protein, is the network of hydrogen bonds throughout the structure. Although the controversy over the relative propensities of these two fundamental elements has already been dealt with in section 1.2.3, it is worth reiterating that this is one of the great unresolved problems in physical-biology. Given the complete freedom a peptide chemist has in choosing a synthetic sequence, his endeavours as much as anyones are predisposed to the study of the structural stability of proteins.

In the previous work it was conclusively demonstrated that it is possible to chemically synthesise and subsequently purify ubiquitin in a biologically active and conformationally authentic form. Having achieved this it was now possible to realize our principle objectives by designing a ubiquitin analogue which would explore the stability of the tertiary structure in the protein. Access to the crystallographic dataset (1UBQ in the Brookhaven database) on an Evans and Sutherland molecular graphics workstation made it possible to study closely the tertiary structure of ubiquitin. Careful inspection of the protein's hydrophobic core revealed it comprised primarily seven residues, the side-chains of which were closely packed together in the heart of the structure (see table 2.9).

	L67	V5	I30	L43	L56	V26	L55
L67							
V5	5.70						
I30	8.53	4.21					
L43	4.48	5.07	4.58				
L56	4.89	8.96	9.13	6.66			
V26	4.70	5.46	4.26	4.01	3.87		
L55	4.18	8.57	8.84	4.82	6.89	7.22	

Table 2.9 Relative Distance (A^o) Between Terminal Methyl Carbon Atoms in the Side-Chains of Hydrophobic Core Residues

Armed with this information it was decided to investigate the importance of the hydrophobic core in stabilizing the tertiary structure of ubiquitin. To that end the seven key residues in the core were replaced by the small non-hydrophobic amino acids shown in figure 2.32.

Val5 to ^{15}N-Glycine
Val26 to Alanine
Ile30 to Alanine
Leu43 to ^{15}N-Glycine
Leu50 to ^{15}N-Glycine
Leu56 to ^{15}N-Glycine
Leu67 to ^{15}N-Glycine

Figure 2.32 Substitutions in the Hydrophobic Core of Ubiquitin

Glycine was used to replace 5 of the 7 amino acids, essentially because it has no side-chain and is thus relatively non-polar. Isotopic enrichment of these additional glycines with ^{15}N was intended to allow easier identification in an NMR experiment. For example, a ^{1}H/^{15}N heteronuclear correlation spectrum of the analogue would clearly distinguish these five residues from the remainder of the proton resonances in the protein [212]. This is extremely useful, since these ^{15}N-glycine handles could then be used as starting points for sequential assignment of the rest of the ^{1}H resonances in the spectrum [121].

Valine-26 and isoleucine-30 are both involved in the α-helix in ubiquitin which incorporates residues 23-34. Given that we had no desire to deliberately destroy this element of secondary structure, these residues had to be replaced with something able to propagate the helix whilst being non-hydrophobic. The work of Baldwin[213] and recently Fersht [214], indicates glycine is unable to sustain an α-helix when located at an internal position (as in this case). However, in the same location alanine appears to better accommodate a helical structure. Fortunately, alanine possesses a relatively small side-chain, resulting in it being the obvious choice for substitutions in this area of ubiquitin.

These replacements amounted to the removal of 24 methylene groups, which according to Fersht [141,142] should represent a loss in stabilization of between 24 and 38 kcal mol^{-1}. We have already calculated the free energy of stabilization of ubiquitin to be about 8.0 kcal mol^{-1}, therefore the above loss in free energy would be enough to completely devastate the compound's tertiary structure. Indeed, replacement of Leu-67 and Leu-69 with Asn is known to completely eliminate ubiquitin's ability to conjugate proteins [130].

To recapitulate, what was done was to completely remove the hydrophobic core from this small globular protein, in an attempt to elucidate whether hydrogen bonding alone is sufficient to hold together the native structure.

2.5.2 Synthesis and Purification of Ubdes-Core

The experience drawn from the previous work on the synthesis of ubiquitin was employed in the construction and purification of this ubiquitin analogue. The design of this analogue required a relatively large amount of Fmoc^{15}N-glycine (**53**) starting material. Although commercially available, this material is rather expensive. Fortunately, it is easily prepared from ^{15}N-glycine (**51**) and 9-fluorenylmethyl succimimidyl carbonate (**52**) using the procedure of Shute and Rich [215].

Ubdes-Core contains the same number of amino acids as the natural sequence, therefore a similar synthetic strategy to before was considered appropriate. Initial attachment of Gly-76 to the Wang support via the amino acid chloride gave a sufficiently low loading (40.5%, 0.291 mmole/g) to allow unhindered elongation of the resin bound peptide chains. As in the previous synthesis, the remaining 60% of

the functional groups on the support were capped with benzoyl chloride and the loaded resin was checked for any glycylglycine content, none of which was found.

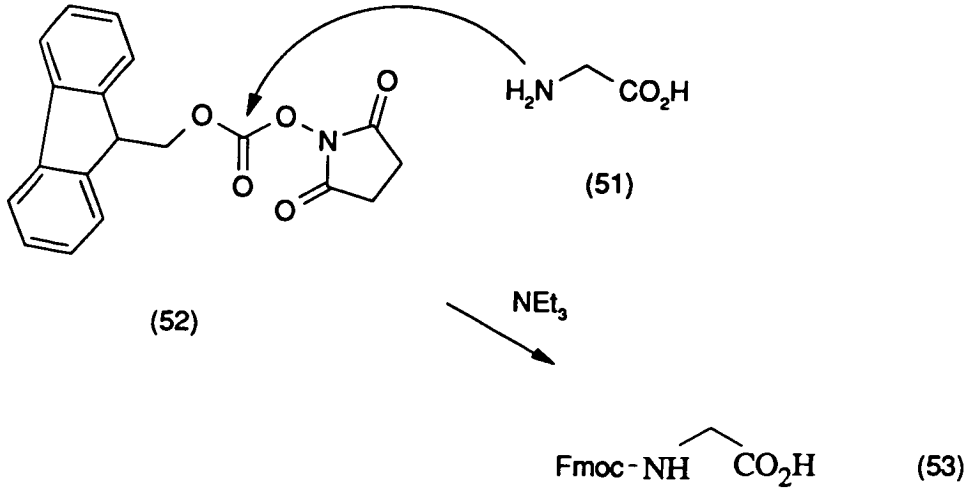


Figure 2.33 Preparation of Fmoc Amino Acids

The synthesis was again carried out on an ABI 430A machine (0.25 mmole scale) using 1:1 DMF:dioxan as the solvent during all coupling steps. Most amino acids, with the same exceptions as before, were triple coupled and any unreacted amine groups on the peptide capped with acetic anhydride at the end of every cycle. Throughout the synthesis all the piperidine initiated Fmoc deprotections were monitored using the real-time UV system discussed in section 2.4.1 (see figure 2.34).

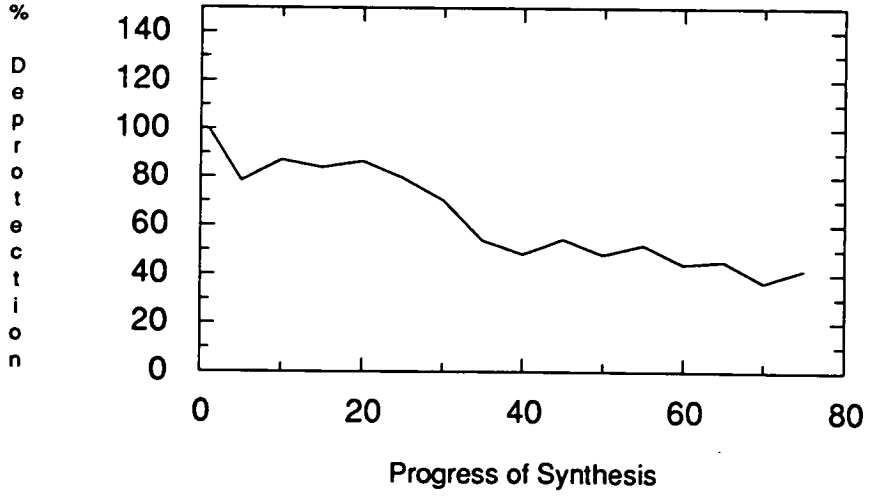


Figure 2.34 Deprotection Profile for Ubdes-Core

The overall efficiency of the synthesis on the resin was 44%, as judged by the UV substitution of the final methionine relative to the theoretical value calculated for 100% yield on every step of the synthesis.

A trial cleavage was carried out on a small sample of the resin and as before anisole, thioanisole and ethyl methyl sulphide were found to be the best scavengers. The optimum time of 4.5 hours was slightly longer than the 3 hours required to cleave the wild-type sequence. Electrospray mass spectrometry [195] was performed on this crude material and clearly showed the correct molecular ion (expected mass=8234.2 Da) was present, as well as other truncated and partially protected species (see figure 2.35).

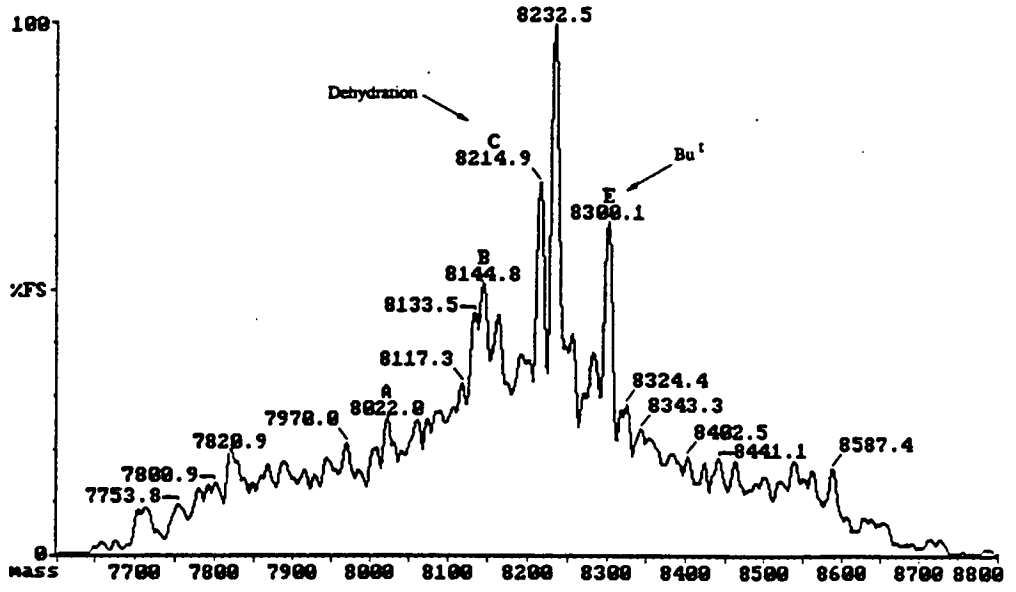


Figure 2.35 Electrospray Mass Spectrometry on Crude Ubdes-Core Showing the Correct Molecular Ion at 8232.5 Da

The cleavage conditions now having been established, the bulk of the resin was divided into three equal portions and each was cleaved and purified in the same way. An almost identical purification protocol to before was employed, namely:

1. **Deprotection** - 20% piperidine in DMF, 3 x 10 minutes.

2. **Cleavage** - 4.5 hours under nitrogen, 80% TFA and 5% each of H_2O, anisole, thioanisole and ethyl methyl sulphide.

3. **Sephadex G50 Gel Filtration** - Eluting buffer contained 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5.

4. **Stepwise Dialysis** - stepwise removal of urea and finally DTT.

5. **CM-Sepharose Cation Exchange** - pH gradient of 4.5 to 5.5, 50 mM NH_4OAc followed by a salt gradient of 50 mM to 0.3 M NH_4OAc at pH 5.5.

6. **DEAE-Sepharose Anion Exchange** - salt gradient of 50 mM to 0.3 M NH_4OAc at pH 9.30.

These were followed by one additional step.

7. **Semi-Preparative HPLC** - C_8 RP300 column, gradient system A.

The crude polypeptide material behaved more or less as expected in all steps up to the DEAE-Sepharose column where difficulties were experienced in attempting to bind the protein to the anion exchange matrix. Indeed, the material washed straight through on several occasions. These problems were eventually overcome by making sure the protein solution was exactly (within 0.05 of a pH unit) the same pH as the column and also by running the column in starting buffer for long periods of time before attempting a salt gradient. Such a difficulty had not been encountered when working with synthetic or natural ubiquitin, and it certainly could not have been predicted from isoelectric focussing since Ubdes-Core has a pI identical to the wild-type sequence (see figure 2.38). The most likely explanation for this observation is that Ubdes-Core is extremely unstable at this basic pH. Even if the analogue had folded into a tertiary conformation it was very unlikely that this structure would be resilient to such a high pH. Therefore, the conditions employed in this step would probably be enough to denature the analogue and thus drastically alter its ion exchange behaviour.

The problems associated with anion exchange meant that it was usually necessary to carry out an additional semi-preparative HPLC step before homogeneous material could be isolated (see figure 2.36).

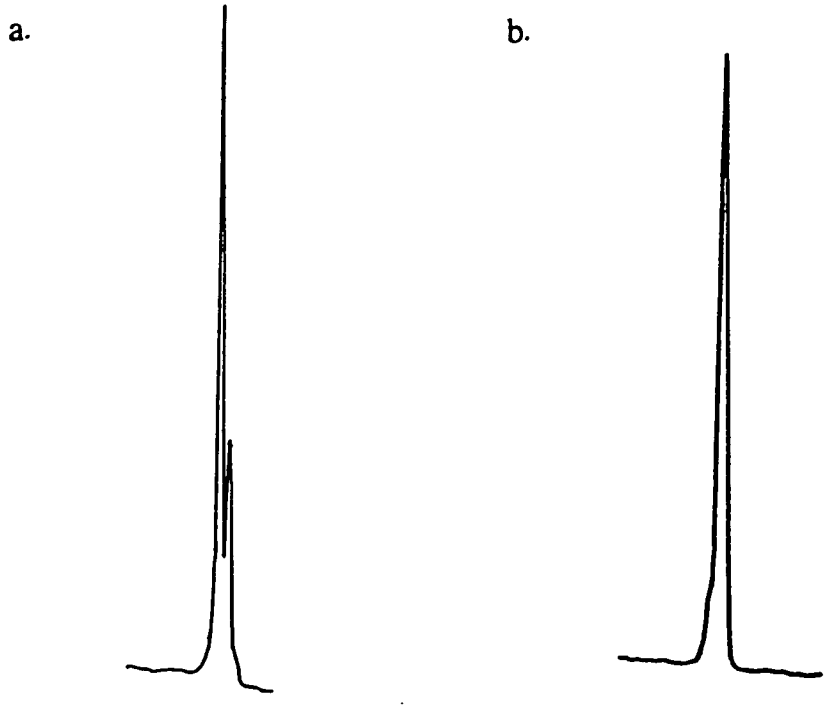


Figure 2.36 HPLC Trace of Ubdes-Core Before (a) and After (b) HPLC Purification. C_8 RP300 column, gradient system A

Minor purification problems aside, the overall yield of Ubdes-Core was a respectable 81 mg or 3.9%. Compared with the 40% yield on the resin this figure does indicate that a considerable amount of material was lost during cleavage, and to a lesser extent purification of the ubiquitin analogue.

Exhaustive characterization of this material was carried out in a manner analogues to that of synthetic ubiquitin. Amino acid analysis, carried out after each stage of the purification, served as a good indicator of the increasing homogeneity of the polypeptide (see table 2.10). Acid and enzymic hydrolyses of the purified material confirmed that it had an amino acid composition consistent with that of Ubdes-Core, and that histidine racemisation and β-aspartyl formation had once again been avoided.

Amino Acid	Expected	Gel Filtration	CM Sepharose	DEAE Sepharose	Enzymatic Digest [a]
Asp/Asn	7 (5)	6.94	7.06	7.06	4.71
Thr	7	5.02	5.59	6.22	9.70
Ser	3	2.24	2.27	2.23	4.97
Glu/Gln	12 (6)	12.30	12.96	12.60	6.97
Pro	3	2.33	2.60	2.76	2.41
Gly	11	12.46	12.18	11.15	14.80
Ala	4	3.81	4.02	4.00	3.60
Val	2	1.89	2.02	2.00	2.79
Met	1	0.56	0.78	0.94	Met O
Ile	6	5.12	5.57	5.85	5.92
Leu	5	4.95	5.13	4.97	6.34
Tyr	1	0.99	0.97	0.90	1.18
Phe	2	1.59	1.75	2.05	2.33
His	1	1.14	1.12	0.97	0.98
Lys	7	6.30	6.84	7.05	6.58
Arg	4	4.04	4.19	3.70	3.99

Table 2.10 Acid and Enzymatic Hydrolyses of Ubdes-Core
[a] Pepsin, Prolidase and Aminopeptidase.
Figures in Parenthesis are Asp and Glu Contents

Enzymatic digestion with trypsin revealed the isolated material had a primary sequence consistent with that of Ubdes-Core. Isolation of the resulting peptide fragments by HPLC followed by amino acid analysis of these, confirmed the tryptic map was that of the analogue.

Electrospray mass spectrometry indicated the isolated material had a mass of 8234.0 Da which compares favourably with the theoretical mass of 8234.2 Da calculated for Ubdes-Core [195]. As well as further establishing the authenticity of the polypeptide, electrospray mass spectrometry provided additional evidence for its high level of purity (see figure 2.37).

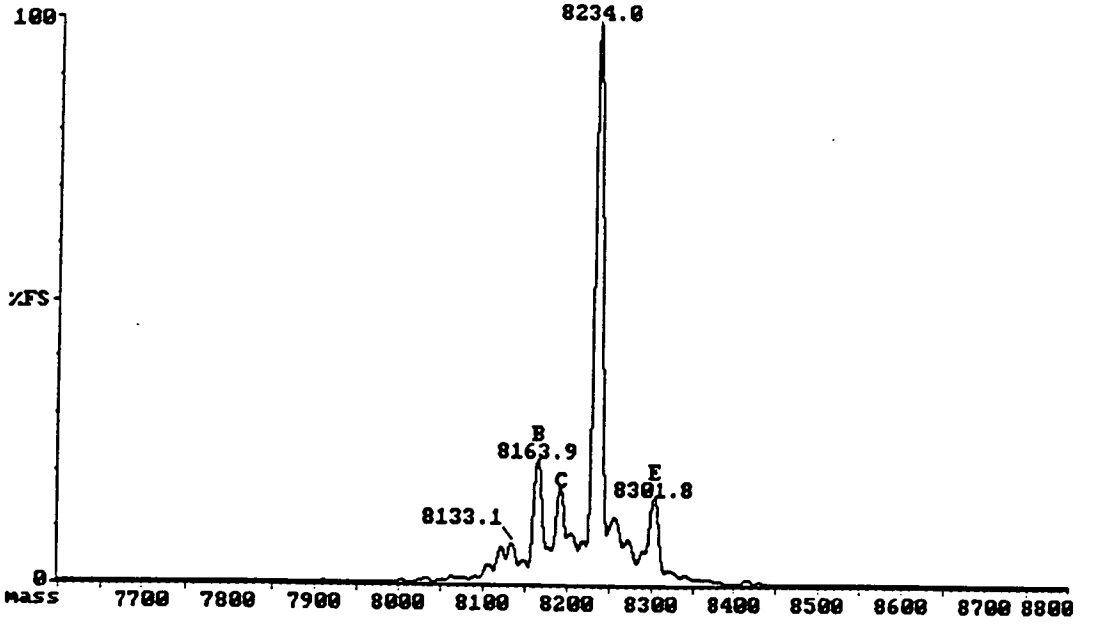


Figure 2.37 Electrospray Mass Spectrometry on Ubdes-Core Showing the Calculated Mass of 8234.0 Da

The isoelectric focussing results shown in figure 2.38 neatly illustrate how successful ion exchange chromatography can be in purifying a polypeptide of this size. Anion exchange chromatography was, in spite of the afore mentioned problems, clearly able to resolve the three major components that comprised the principle fraction from cation exchange (lane8). Furthermore, lane3 not only reveals Ubdes-Core to have an identical pI to ubiquitin, but also provides additional confirmation of its high degree of homogeneity.

Ubdes-Core was found to be recognized by anti-ubiquitin IgG polyclonal antibodies in a Western blot assay similar to the one performed on synthetic ubiquitin. While this was certainly an encouraging observation, it was considered dangerous to read too much into it at this stage. The polyclonal nature of these antibodies probably means that they would recognize many of the ubiquitin-like epitopes remaining in the analogue. Many of these would of course be antigenic regardless of the overall structure of the analogue. Therefore, the result can in no way be regarded as evidence for Ubdes-Core having a native ubiquitin conformation. A far better antibody would be a monoclonal antibody raised to a discontinuous epitope on the surface of the native conformation of ubiquitin.

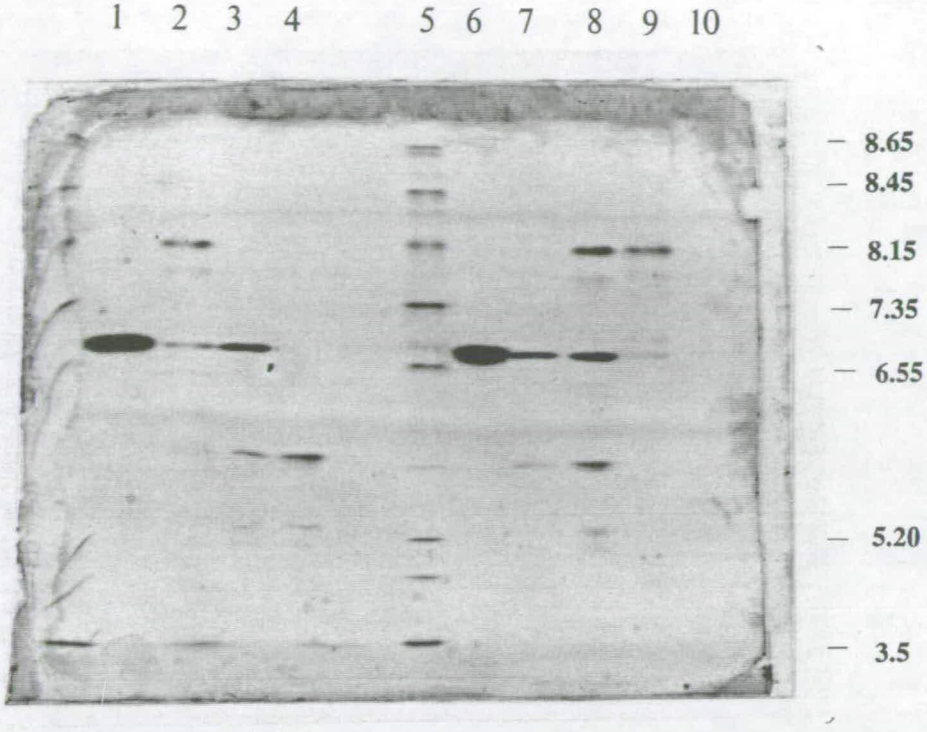


Figure 2.38 **Isoelectric Focussing Gel of Ubdes-Core at Various Stages of Purification**

lane1 = Bov Ub	**lane6 = Bov Ub**
lane2 = DEAE F1	**lane7 = CM-Seph. F2**
lane3 = DEAE F2 *	**lane8 = CM-Seph. F3 ***
lane4 = DEAE F3	**lane9 = CM-Seph. F4**
lane5 = Standards	**lane10 = CM-Seph. F5**

*** - fractions isolated for further purification**

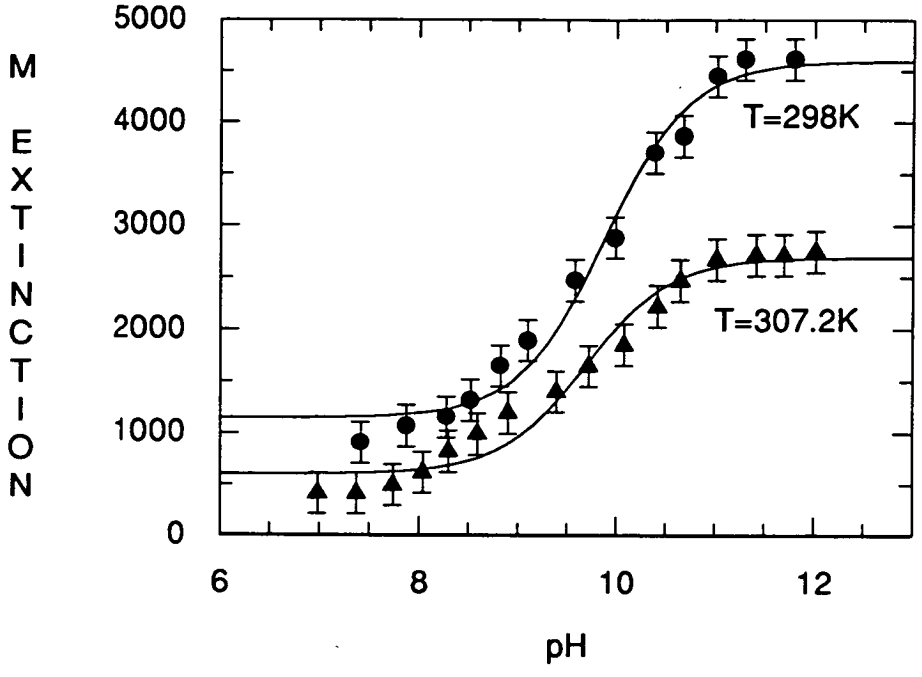
2.5.3 Structural Studies on Ubdes-Core

Having sufficiently characterized the primary structure of Ubdes-Core we were then able to turn our attention to the real objectives of this study. In carrying out this comprehensive structural analysis our approach was, as before, to make use of classical techniques, as well as the more exhaustive NMR and if possible X-ray crystallography techniques.

2.5.3(i) Ultra-Violet Absorption Spectroscopy

Once again the lone tyrosine residue in the protein proved to be invaluable in probing the conformation of Ubdes-Core. Spectrophotometric titration of this

tyrosine revealed its phenolic group had a pKi of 9.88 at 25.0°C and 9.68 at 35.2°C (see graph 2.7).



Graph 2.7 Spectrophotometric Titration (295 nm) of the Phenolic Group of Tyr-59 in Ubdes-Core at Two Temperatures

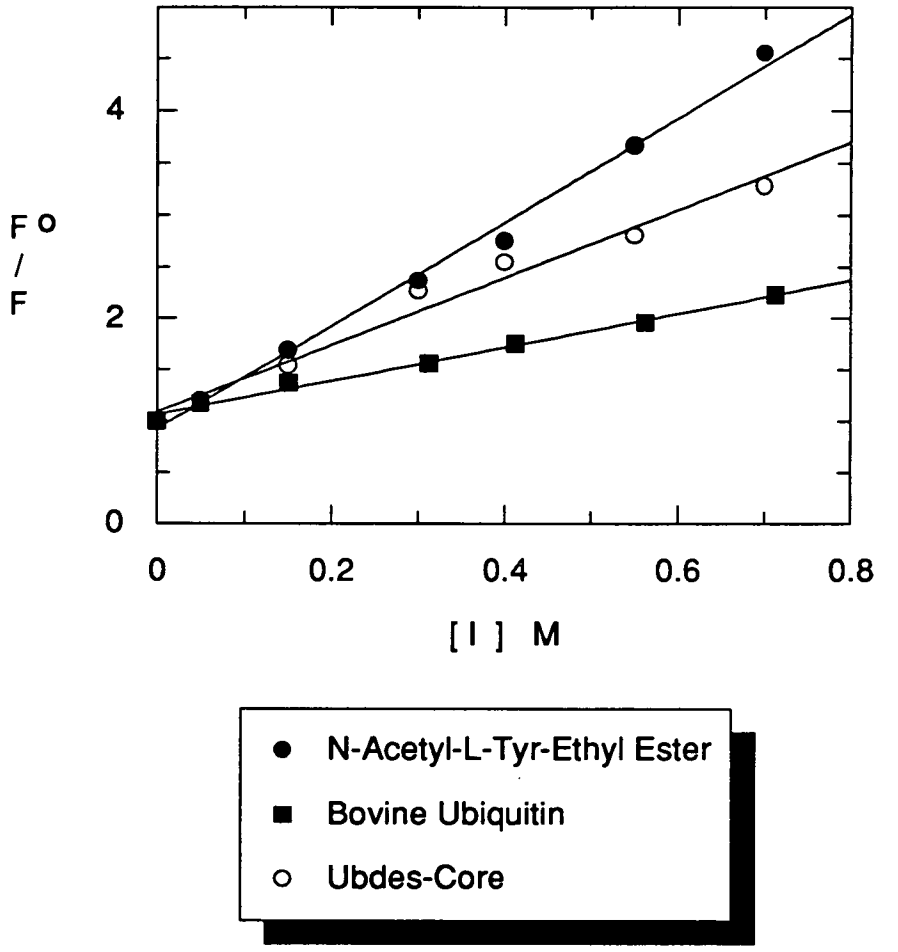
From this data the heat of ionisation (ΔHi) of the phenolic group was calculated to be 9.3 kcal mol^{-1}. This represents a significant drop from the 10.8 kcal mol^{-1} determined for the wild type protein. Evidently removing the hydrophobic core from ubiquitin results in the hydrogen bond environment around Tyr-59 becoming destabilized, hence the lower value of ΔHi for Ubdes-Core. It is however worth pointing out that this value of ΔHi is still substantially higher than for many other proteins and for free tyrosine, perhaps suggesting the phenolic group is involved in at least some hydrogen bonding (see table 2.5 for comparisons).

2.5.3(ii) Fluorescence Studies on Ubdes-Core

A similar set of fluorescence based studies were performed on Ubdes-Core as had previously been carried out on bovine and synthetic ubiquitin.

Quenching of the tyrosine fluorescence by addition of potassium iodide can provide valuable information on the degree of exposure of the residue to the bulk solvent.

Analysis of the Stern-Volmer plot obtained for Ubdes-Core reveals its tyrosine is more exposed to the solvent than this fluorophore in purified bovine ubiquitin (see graph 2.8).

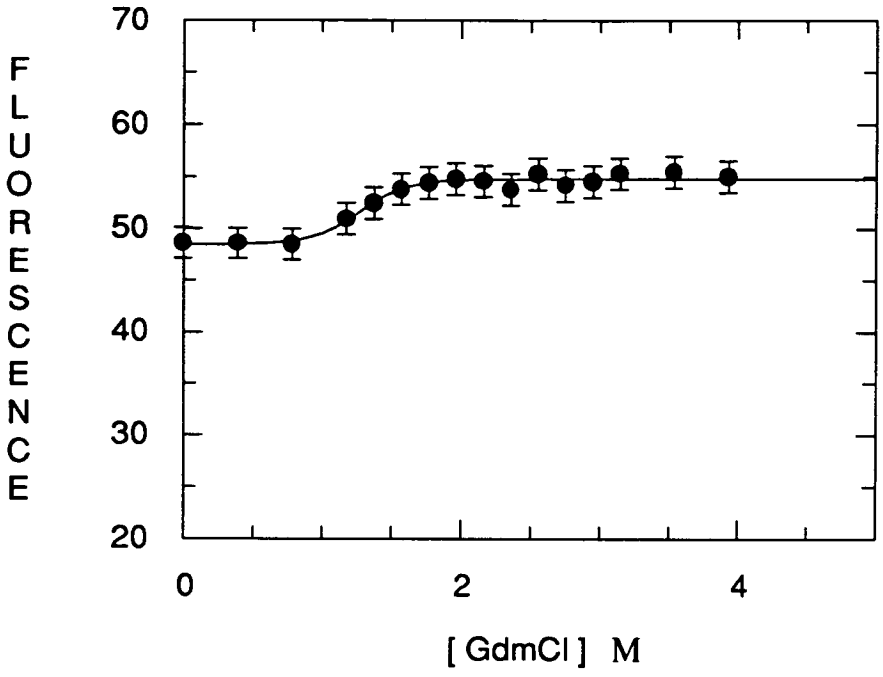


Graph 2.8 Stern-Volmer Plots of Ubdes-Core, Bov Ub and ATEE

A Stern-Volmer constant of 3.27 ± 0.21 M^{-1} signified the tyrosine in Ubdes-Core was not as buried as in natural ubiquitin (K_Q = 1.76 M^{-1}). Nevertheless, a comparison with the N-acetyl tyrosine ethyl ester control (K_Q = 5.00 M^{-1}) clearly indicated some shielding from the bulk solvent.

Monitoring the denaturation of Ubdes-Core by fluorescence led to a value of 4.3 ± 1.5 kcal mol^{-1} being obtained for its free energy of stabilization (see graph 2.9). Although there are rather large error boundaries associated with this value

(primarily due to the small difference between F_N and F_U) it is clearly much smaller than the ΔG_{H2O} calculated for bovine ubiquitin (8.0 kcal mol^{-1}).



Graph 2.9 GdmCl Induced Unfolding of Ubdes-Core

A value of around 3.7 kcal mol^{-1} for $\Delta\Delta G_{H2O}$ (the change in ΔG_{H2O} between wild type and analogue) does confirm the importance of the hydrophobic effect. However, it is somewhat smaller than the huge loss in stability predicted by the Fersht model [141, 142]. It was also interesting that deprived of the hydrophobic effect ubiquitin still had approximately 4.3 kcal mol^{-1} nett stabilization from hydrogen bonding contributions. Whether this was enough to maintain the ubiquitin fold remained to be seen.

2.5.3(iii) Circular Dichroism Studies on Ubdes-Core

Circular dichroism was again performed in Stirling University [204]. CONTIN secondary structure analysis significantly revealed there to be no α-helix in Ubdes-Core; this was in spite of the efforts made to keep the helix intact (see figure 2.39). In addition, this analysis indicated the β-sheet content in Ubdes-Core was approximately the same as in natural ubiquitin (see table 2.11). It is noteworthy that while Ubdes-Core apparently contains a similar amount of β-sheet as the natural sequence, it cannot be determined from CD whether these sheets bear any resemblance to one another in a qualitative sense. Loss of the helix may be due to

elimination of important tertiary interactions between core residues and this element of secondary structure. Indeed, both valine-5 and leucine-43 do lie very close to the helix in the natural fold. Furthermore, the ubiquitin fragment Ub1-35 which contains the α-helix has been found to have no structure in solution by NMR spectroscopy [216], suggesting that the remainder of the ubiquitin structure is required to stabilize the α-helix. Of course it is also possible that the two substitutions made in the helix itself led to its destabilization in Ubdes-Core. The ubiquitin analogue in which these two residues have been replaced with the wild type amino acids is currently being synthesised, and should provide important information with regard to the stability of the helix. It is perhaps interesting to note that secondary structure prediction programs (Garnier-Osguthorpe-Robson [217] and Chou-Fasman [218]) intimated that the α-helix would be retained in Ubdes-Core and that the β-sheet would be reduced.

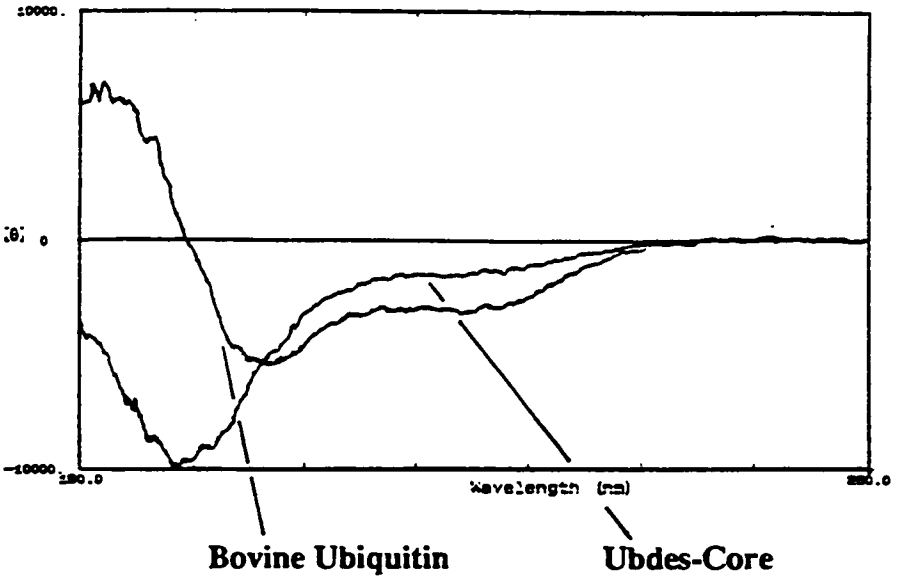


Figure 2.39 Far UV CD Spectra of Ubdes-Core and Purified Bovine Ubiquitin in 50 mM Boric Acid at pH 7.0

Previous studies by Wilkinson and Mayer had indicated increased helicity could be induced into the ubiquitin fold by exposing the protein to various alcohols [131]. Upon dissolving Ubdes-Core in 50% aqueous trifluoroethanol (TFE) the analogue was observed to undergo an intriguing structural reorganization as judged by circular dichroism spectroscopy (see table 2.11). Indeed the secondary structure content of Ubdes-Core in 50% TFE was remarkably similar to bovine ubiquitin dissolved in water, and although a qualitative description of this alcohol generated

conformation was impossible from this CD spectrum a ubiquitin like fold could not be ruled out. Bovine ubiquitin was itself subject to a substantial structural rearrangement upon exposure to TFE (see table 2.11). This observation is in keeping with the original results of Wilkinson and Mayer.

Example	α-helix	β-sheet	Remainder
Bov Ub in H_2O	15 ± 2.5%	57 ± 2.6%	28 ± 4.5%
Ubdes-Core in H_2O	0 ± 3.4 x 10^{-7}%	52 ± 1.5%	48 ± 1.5%
Bov Ub in 50% TFE	63 ± 0.73%	37 ± 0.73%	0 ± 7.1 x 10^{-7}%
Ubdes-Core in 50% TFE	17 ± 1.0%	55 ± 1.1%	28 ± 1.9%

Table 2.11 Secondary Structure Contents derived from CD using CONTIN Procedure

2.5.3(iv) X-Ray Crystallography

Along with NMR spectroscopy, X-ray crystallography provides unrestricted information on the tertiary structure of a protein and was thus a technique we wished to make use of in this study. In the previous work we were able to use natural ubiquitin crystals to initiate the crystallization of the synthetic material, however, this approach could not be extended to the crystallization of the chemically different Ubdes-Core. Therefore, attempts were made at the *ab initio* crystallization of this analogue and while 20 different vapour diffusion experiments were set up, no crystals could be grown in the following three month period. Protein crystallization is generally accepted as being a rather unpredictable affair and it is not inconceivable that the required conditions were not covered by our scan. Equally a protein devoid of any organized tertiary structure will not crystallize for entropic reasons, perhaps providing an alternative explanation for our lack of success.

2.5.3(v) Nuclear Magnetic Resonance Studies on Ubdes-Core

Two dimensional NMR was used to investigate in detail the solution conformation of Ubdes-Core. Comparison of the TOCSY spectrum obtained for the analogue

with that acquired for purified bovine ubiquitin very quickly confirmed that removal of the hydrophobic core from ubiquitin prevents the protein from folding properly, if at all (see figure 2.40).

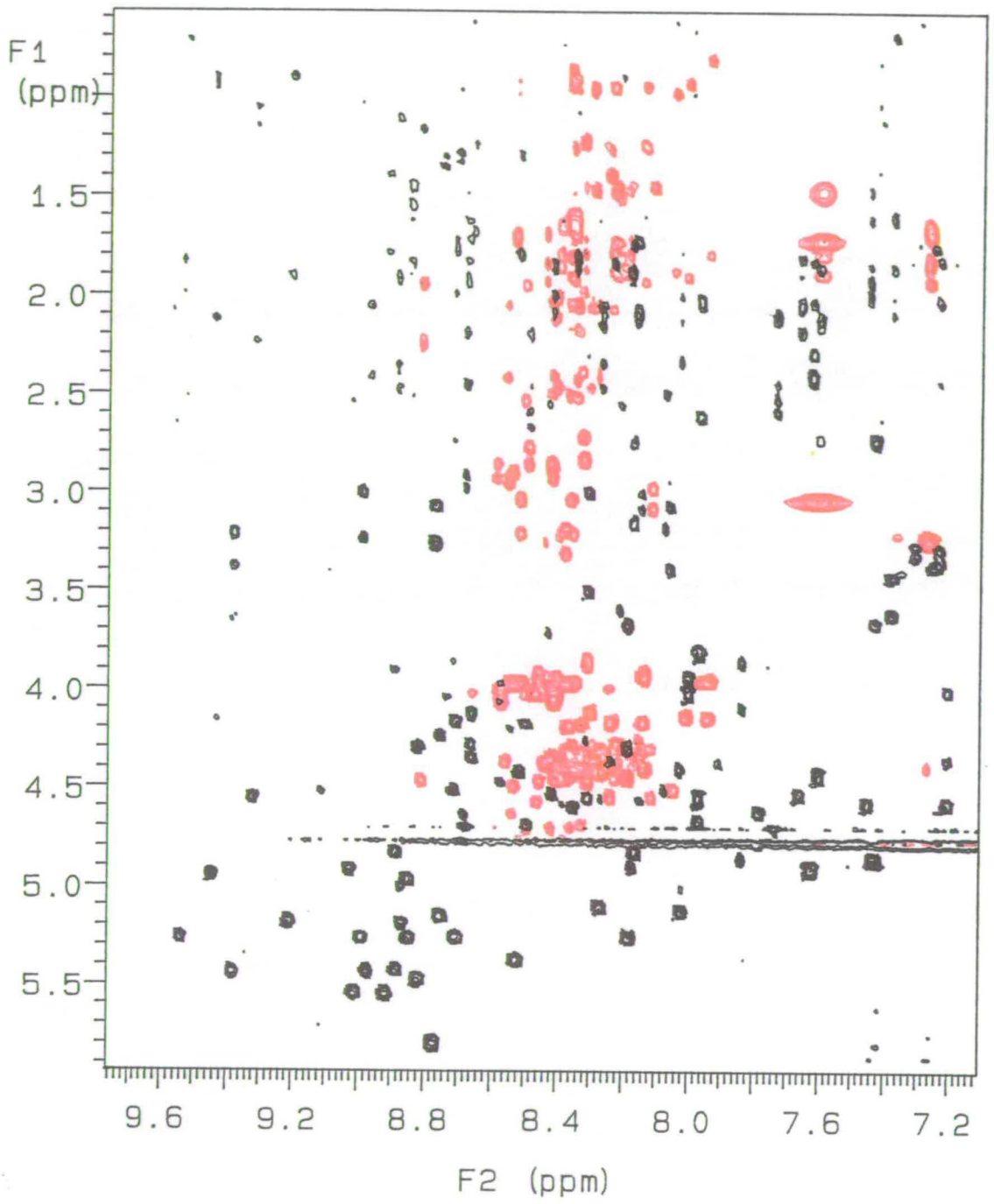


Figure 2.40 **Region from the TOCSY Spectra of Ubdes-Core (red) and Purified Bovine Ubiquitin (black) obtained in 90% H_2O/ 10% D_2O at pH3.0 and 50°C**

The lack of crosspeak dispersion in the TOCSY spectrum of the analogue was in stark contrast to the large distribution in the TOCSY spectrum of bovine ubiquitin. Given the sensitivity of this region to the solution conformation of a protein, this observation was considered quite conclusive evidence that Ubdes-Core did not have the native ubiquitin fold.

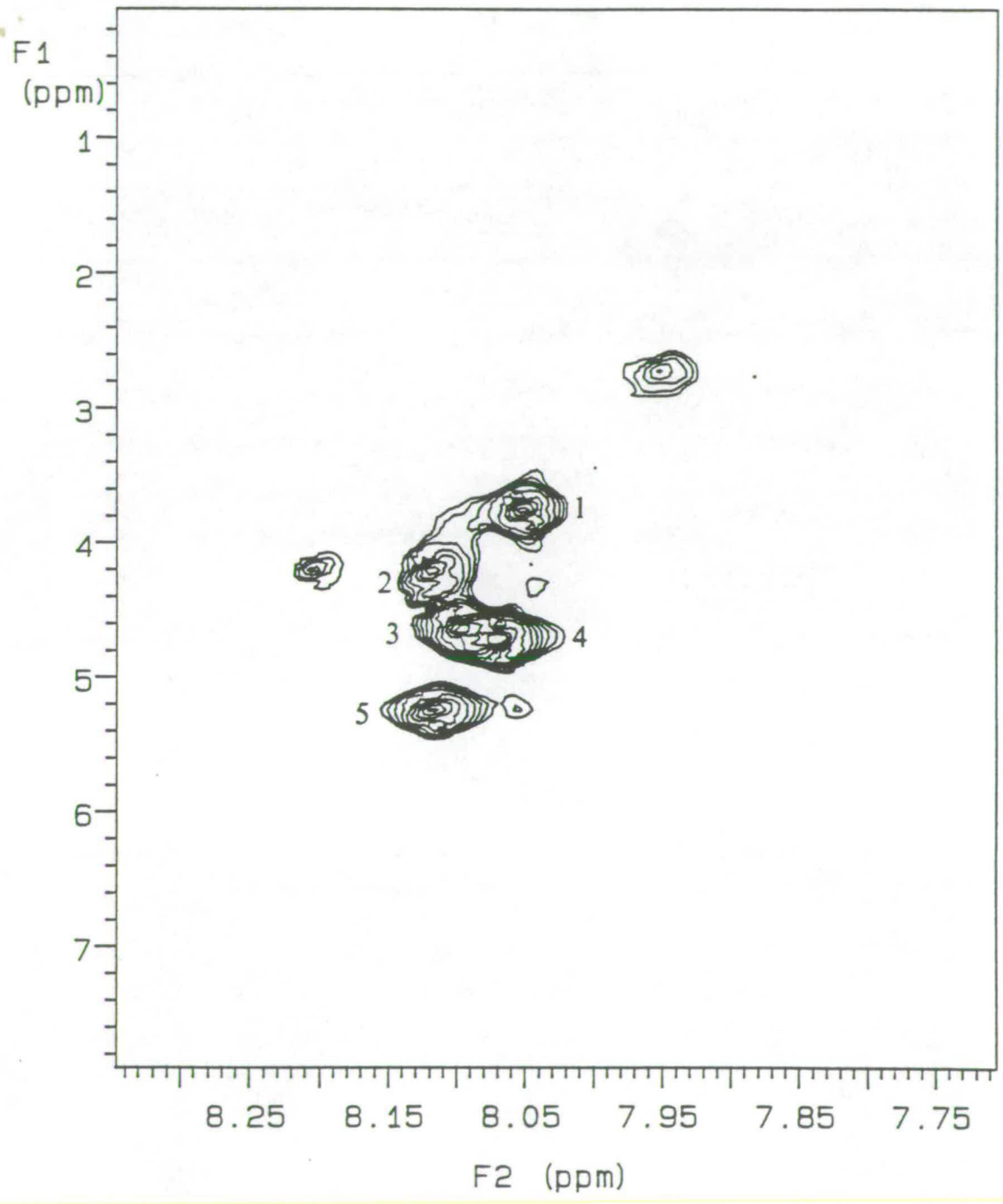


Figure 2.41 Heteronuclear Multiple Quantum Coherence Transfer Spectrum of Ubdes-Core showing the 5 labelled Glycines

Although the isotopic enrichment of the five glycine substitutions with ^{15}N did allow unambiguous identification of their proton resonances (see figure 2.41), the remainder of the proton resonances were impossible to assign due to serious spectral crowding and overlap.

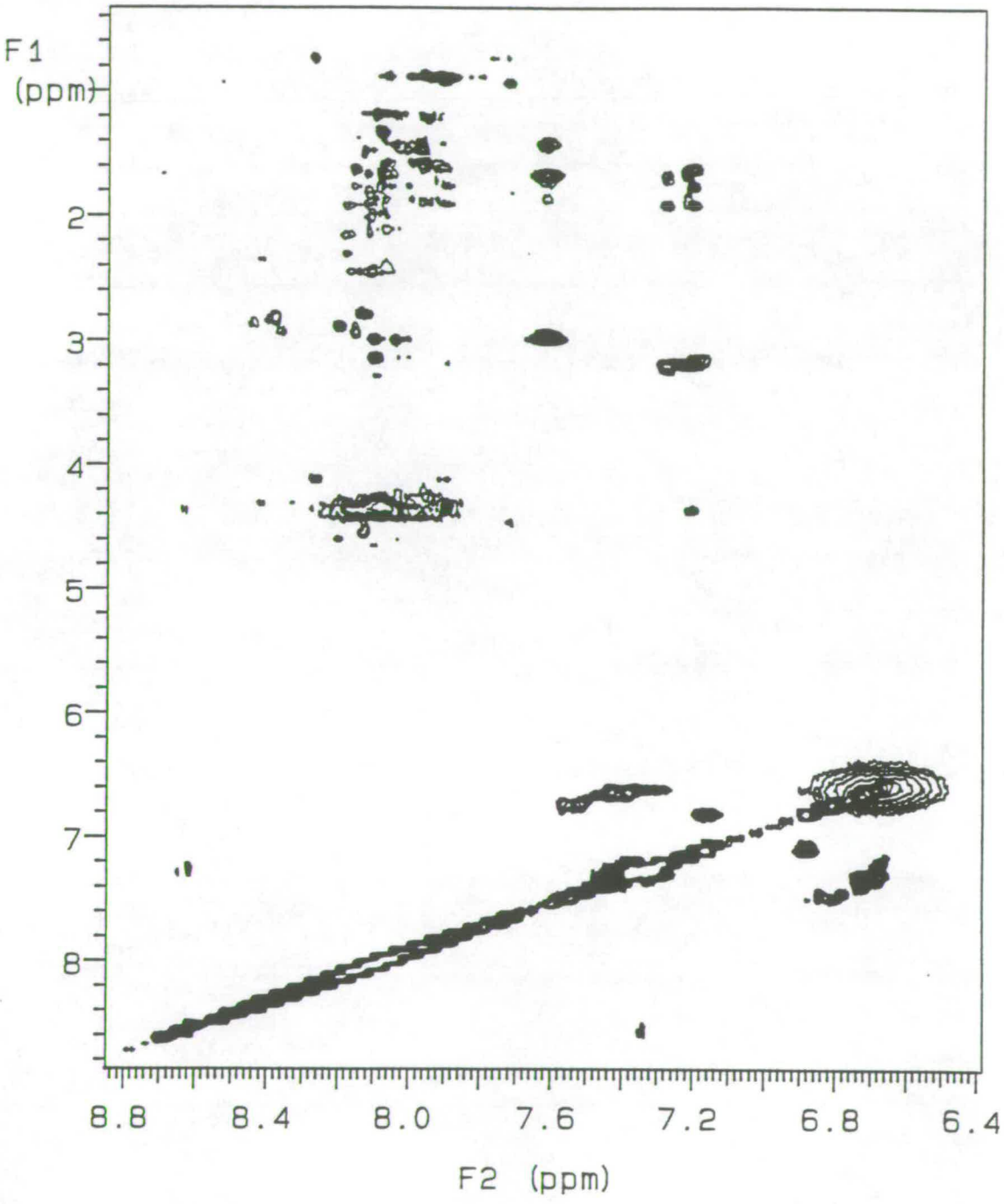


Figure 2.42 Region from TOCSY Spectrum of Ubdes-Core dissolved 50% TFE / 50% 9:1 H_2O : D_2O at pH 6.0

The analogue apparently gives rise to a form of collapsed 2D NMR spectrum usually indicative of proteins in which there is no well defined tertiary structure. This finding is seemingly at odds with the UV, CD and fluorescence studies all of which suggest the analogue does have some detectable structure. However, in a recent paper Tsai and co-workers reported that a phospholipase A_2 mutant exhibited secondary structure when examined by CD spectroscopy, but failed to show any amide chemical shift dispersion in a 1D NMR experiment [219]. This group suggested the mutant exists in a molten globule state containing isolated elements of secondary structure, but no stabilizing tertiary structure interactions. Consequently, the mutant has many internal degrees of freedom which result in this secondary structure being invisible on an NMR timescale (ms). It is therefore possible that Ubdes-Core possesses elements of secondary structure such as β-sheet which are detectable on an electronic timescale, but cannot be observed on the much longer NMR timescale. Indeed, circular dichroism spectroscopy clearly indicates that Ubdes-Core contains more secondary structure when dissolved in 50% TFE, whereas the TOCSY spectrum of the analogue in this medium is little different from that of the protein in water (see figure 2.42). The inconsistent nature of these observations points to NMR not being able to detect elements of secondary structure in dynamic equilibrium.

2.5.4 Ubiquitin Thioester and Protein Conjugation Activities of Ubdes-Core

In a further collaboration with Mayer and co-workers in Nottingham the ability of Ubdes-Core to form an E1 thioester and therefore endogenous protein conjugates was assayed [210]. Purified bovine ubiquitin and Ubdes-Core were initially radiolabelled with ^{125}I before being incubated with human placental fraction II (FII) supplemented with ATP for 30 minutes at 37°C. The contents of these two *in vitro* systems were then separated by SDS-PAGE and any radiolabelled species visualized using autoradiography. For comparative purposes a similar assay was performed in which any thioesters produced during the incubation were destroyed by incubation with β-mercaptoethanol at 90°C. This additional step resulted in only endogenous protein conjugates being seen by autoradiography.

Analysis of the bovine ubiquitin assay with and without β-mercaptoethanol (lanes 1 and 5 respectively) reveals the Ub-E1 thioester to be the dark band just above 116 kDa. Clearly no thioester can form with Ubdes-Core (lane 6) and correspondingly the analogue is unable to form endogenous protein conjugates

(lane 2). Studies by Ecker *et al* suggest that E1 recognizes both the tertiary structure of ubiquitin and residues in the tail of the protein [130]. Therefore, the inability of Ubdes-Core to form a thioester with the E1 enzyme presents additional evidence that this analogue does not possess the native ubiquitin fold.

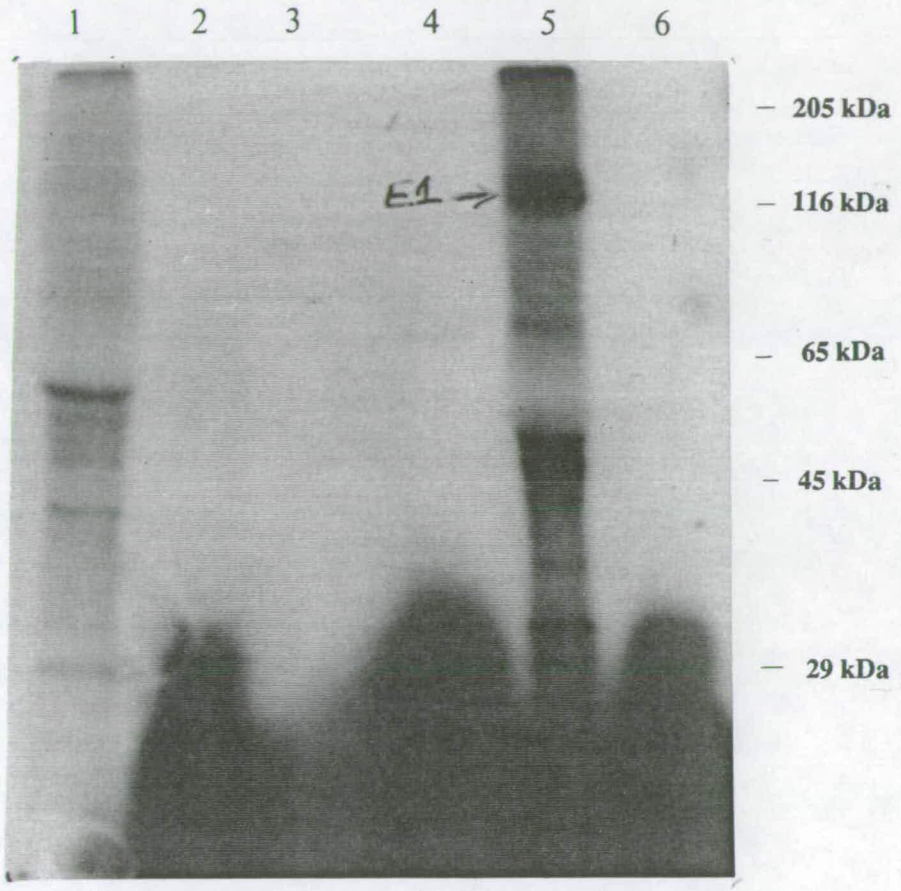


Figure 2.43 **Thioester and Endogenous Protein Conjugation Activities of Ubdes-Core and Bovine Ubiquitin**
Results from Autoradiography
lane1 = Bovine Ub + FII + ATP + β-mercaptoethanol
lane2 = Ubdes-Core + FII + ATP + β-mercaptoethanol
lane3 = Bovine Ub
lane4 = Ubdes-Core
lane5 = Bovine Ub + FII + ATP
lane6 = Ubdes-Core + FII + ATP

2.5.4 Conclusions on the Importance of the Hydophobic Core in Ubiquitin

Removal of the hydrophobic core in ubiquitin appears to have a devastating effect upon the stability and hence structure of the protein. Ubdes-Core cannot adopt the native conformation of ubiquitin, implying that hydrogen bonding contributions are not in themselves sufficient to stabilize such a globular fold. Whether this analogue is able to form an alternative structure is impossible to say with any certainty. Initial UV, CD and fluorescence studies would suggest the analogue contains some structure, however, there seems to be no evidence from NMR to support this. One possibility is that the analogue exists in a molten globule like state in which there are isolated areas of secondary structure.

Section 2.6 Chemical Synthesis of Human CEP52 and Human UbCEP52

2.6.1 Introduction

The recent discovery of a series of genes encoding ubiquitin fused to a carboxy-terminal protein (CEP) has created a whole new area of ubiquitin research. While it is now accepted that CEPs are involved in ribosome biogenesis, there is still relatively little known about the structure of these or of their ubiquitin fused precursors (UbCEPs). All CEPs apparently contain within their primary sequence a putative zinc finger DNA/RNA binding motif. This class of oligonucleotide binding protein is characterized by having a tetrahedrally coordinated Zn^{2+} atom between a region of α-helix and β-sheet [220,221]. There are three general sub-classes of zinc finger, each of which differs in the identity of the four amino acid ligands [102]. Type I zinc fingers, by far the most common, contain a metal centre coordinated to two cysteine and two histidine residues, type II zinc fingers contain four cysteine ligands and type III one histidine and three cysteines. Zinc fingers are so called because the coordinated metal ion holds the α-helix and β-sheet together in such a way as to create a structure capable of forming interactions within the major groove of DNA (see figure 2.44).

Ubiquitin carboxy-terminal extension proteins generally contain sequences which conform to the type II zinc finger sub-class. However, other than sequence similarities there is no direct evidence that these proteins bind zinc or adopt a zinc finger like conformation.

Figure 2.44 A Zinc Finger Interacting within the Major Groove of DNA

To investigate whether CEPs actually behave as zinc fingers and whether this behaviour is in any way modified when the peptide is tethered to ubiquitin, it was decided to chemically synthesise a suitable CEP, as well as its UbCEP precursor. Note that *in vivo* processing of UbCEPs makes isolation of these precursors almost impossible from natural sources. Two CEPs have been identified in the human genome [14,15], a 52 amino acid peptide and an 80 amino acid protein. The latter has a precursor 156 amino acids in length while the former is derived from a 128 amino acid protein. Obviously the construction of UbCEP52 would present less of a problem than the synthesis of the much longer UbCEP80. Therefore, human CEP52 and UbCEP52 were deemed reasonable synthetic targets (see figure 2.45).

H_2N - Ubiquitin - I. I. E. P. S. L. R. Q. L. A. Q. K. Y. N. C. D. K. M.

I. C. R. K. C. Y. A. R. L. H. R. R. A. V. N. C. R.

K. K. K. C. G. H. T. N. N. L. R. P. K. K. K. V. K

-CO_2H

Figure 2.45 Primary Sequence of Human UbCEP52

2.6.2 Solid Phase Peptide Synthesis of CEP52 and UbCEP52

The sequential relationship between CEP52 and UbCEP52 meant that a single linear synthesis could theoretically afford both materials if a portion of the peptidyl-resin was removed at the appropriate point (see figure 2.46).

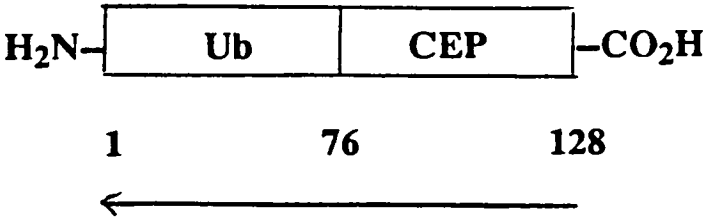


Figure 2.46 The Sequential Synthesis of CEP52 and UbCEP52

All the synthetic considerations which had proved to be so crucial to the success of previous syntheses were reapplied to this synthesis. Initial loading of the support with Lys-128 was achieved by treating the Wang resin with the Fmoc protected amino acid anhydride in the presence of dimethylaminopyridine (DMAP). Unlike the achiral glycine, Fmoc-Lys(Boc)-OH cannot be activated as the amino acid chloride because of the associated racemisation (see figure 1.15). Reducing the reaction time to 15 minutes allowed only 14% (0.1 mmole/g) of the available sites on the resin to be acylated, the remainder were capped with benzoyl chloride. Such an exceptionally low initial loading was considered necessary if interactions between resin bound protein chains were to be minimized.

Solid phase peptide synthesis of initially CEP52 and ultimately UbCEP52 was carried out on an ABI 430A peptide synthesiser (0.1 mmole scale) using a 1:1 mixture of DMF/dioxan as the solvent during all coupling steps. In an effort to obtain near quantitative acylation, most amino acids were coupled once as their symmetrical anhydride followed by twice as their HOBt active esters. Exceptions to this were Fmoc-Gly which was coupled once as a symmetrical anhydride and Fmoc-His(Trt), Fmoc-Asn(Mbh) and Fmoc-Gln(Mbh) all of which were coupled three times as their HOBt active esters. Any unreacted amine groups were subsequently capped with acetic anhydride. Where necessary, amino acid side-chains were protected with the same groups used in the previous two syntheses. All

five cysteine sulphydryls in CEP52 were protected with the reductively removable t-butyl sulphenyl (S-Bu^t) group. The acid and base stability of this protecting group meant that the cysteine residues could be deprotected at a much later stage in the purification, thereby delaying the difficulties associated with handling cysteine containing peptides. Glycine-76, which lies at the interface of Ub and CEP52, was isotopically labelled with ^{15}N. This measure was intended to allow straight forward NMR identification of this strategically important residue which was ultimately hoped to be used as the starting point for sequential assignment in either direction. Real time UV monitoring of the synthesis produced a rather unexpected deprotection profile (see figure 2.47).

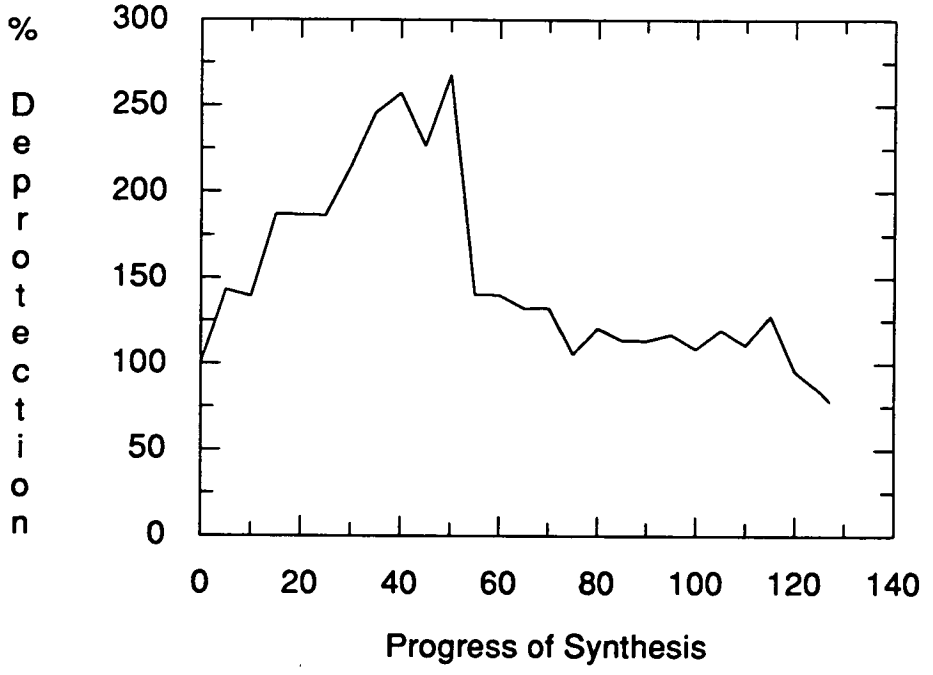


Figure 2.47 Deprotection Profile of the CEP52 and UbCEP52 Synthesis

During construction of the initial CEP52 segment the apparent yield of each deprotection step gradually increased to around 250% of the original deprotection. After 52 cycles approximately one third of the resin was removed, hence the large drop at this point, and thereafter the yields began to decrease in the familiar way. A possible explanation for the initial increase would be a change in the swelling properties of the peptidyl-resin during the first part of the synthesis. This could conceivably result in the deprotection aliquot used in the monitoring having a higher piperidine-fulvene adduct concentration. Manual deprotection of the final

Fmoc protected amino acid, methionine, followed by quantitative UV analysis of this adduct at 302 nm, indicated the overall yield of the synthesis on the resin was 73.5%. If accurate, this value represents an average yield of 99.9% for each of the 254 steps in the synthesis of UbCEP52.

2.6.3 Purification and Characterization of CEP52

In purifying this 52 amino acid peptide a conscious effort was made to leave the five cysteines protected for as long as possible. Therefore, in designing the cleavage conditions care had to taken not to include any scavengers like thioanisole, known to deprotect the reductively removed S-Bu^t group. Preliminary studies revealed a 6 hour aqueous TFA cleavage using ethanedithiol (EDT), EMS, anisole and phenol scavengers to be optimum (see figure 2.49a). Following cleavage the crude peptide was subjected to gel filtration on a Sephadex G50 column using 30% acetic acid as the eluting buffer (figure 2..49b). The material was then gradually dialysed back into water using a 2000 MWCO bag. Reductive removal of the five cysteine protecting groups was effected by treatment with tri-butylphosphine (**54**) in the presence of a small amount of water (see figure 2.48). This process was monitored by HPLC and appeared to go to completion in 3 hours.

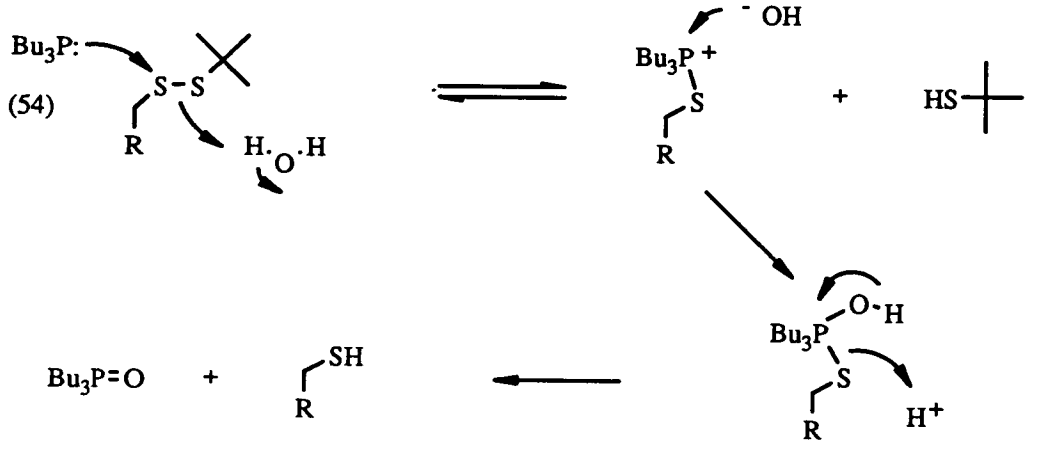


Figure 2.48 Proposed Mechanism for Reductive S-Bu^t Deprotection

Deprotected CEP52 was then exposed to a further Sephadex G50 gel filtration column, this time eluting with a buffer containing 10 mM DTT and 50 mM NH_4OAc at pH 6.0 (see figure 2.49c). Inclusion of the DTT was required to prevent unwanted oxidation of the deprotected cysteines to disulphides. The next

step in the purification, semi-preparative HPLC, was carried out on a C8-RP300 column (gradient system B) and yielded 21mg of essentially homogeneous material (see figure 2.49d).

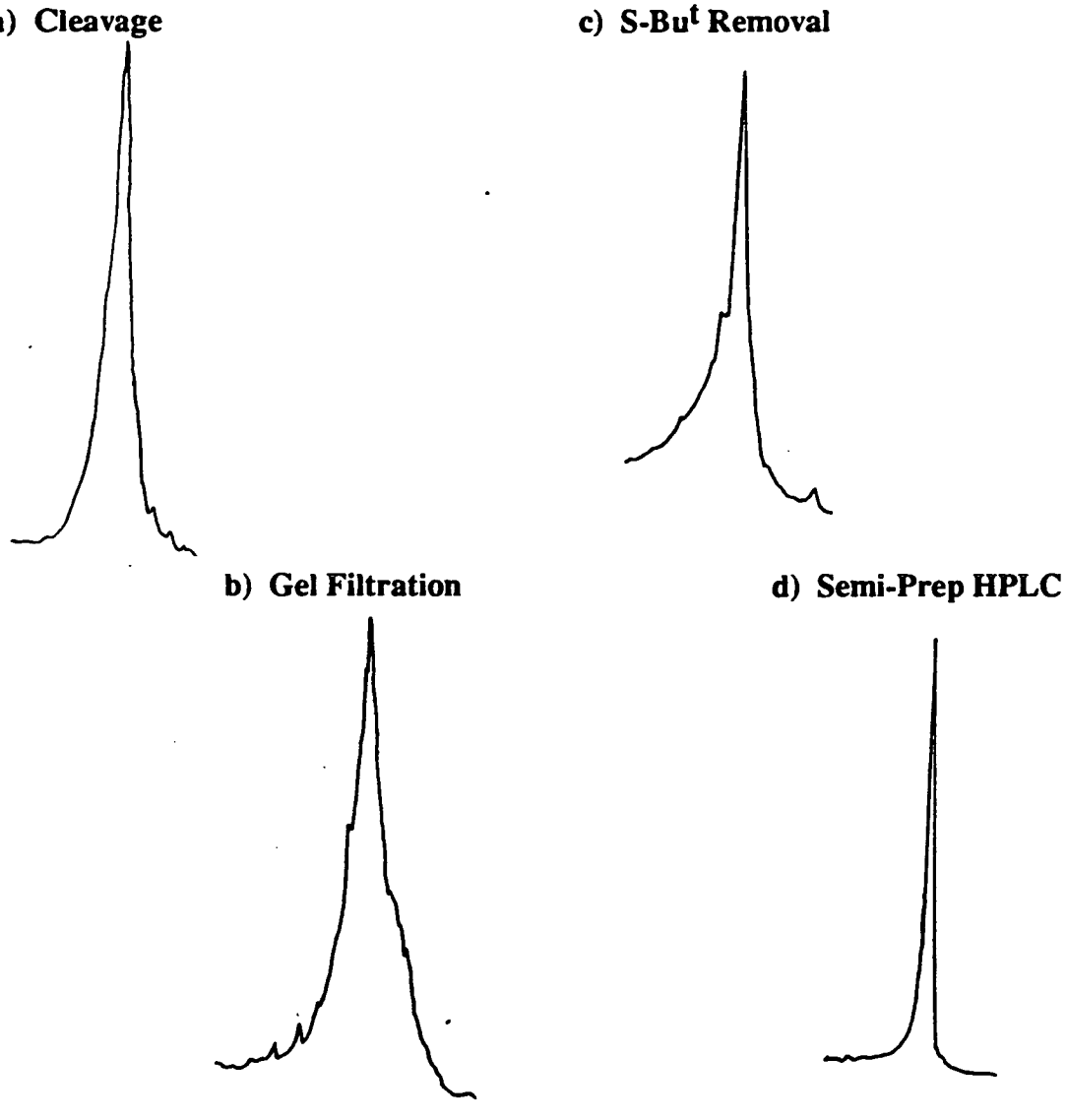


Figure 2.49 **HPLC Analysis of CEP52 at each stage of the Purification C_8-RP300 column, gradient system A**

Amino acid analysis of the purified polypeptide revealed it contained an amino acid composition consistent with the required material (see table 2.12). The low cysteine content was presumably due to oxidation of these residues to cysteic acid. Indeed, this would account for the apparent high proline content since cysteic acid co-elutes with this amino acid.

Amino Acid	Expected	Obtained
Asx	5	4.98
Thr	1	0.73
Ser	1	1.05
Glx	3	4.41
Pro	3	5.61
Gly	1	0.96
Ala	3	3.09
Cys	5	1.41
Val	2	2.00
Met	1	1.17
Ile	3	2.75
Leu	4	4.46
Tyr	2	2.08
His	2	2.31
Lys	10	10.43
Arg	6	6.43

Table 2.12 Amino Acid Analysis of Synthetic CEP52
40 hour hydrolysis in 6N HCl (5mg/ml Na_2SO_3) at 110°C

Electrospray mass spectrometry [195] confirmed the identity of the material to be CEP52 (see figure 2.50). The experimentally determined mass of 6178.8 Da was almost exactly 2 mass units less than the expected value for CEP52 (6180.52 Da). Such a difference can be accounted for by the formation of a single disulphide bridge in the peptide.

Elapsed Time	No of Free Thiols
Immediately after Reduction	4.72
1 hour	2.50
24 hours	1.14

Table 2.13 Oxidation of Synthetic CEP52 in Water

As mass spectrometry had suggested the cysteines in the peptide were prone to oxidation, this possibility was further investigated by dissolving the peptide in water and periodically determining the number of free thiols present using a standard Ellmans assay [222]. Table 2.13 summarises the results which indicate one disulphide bond forms after about an hour and a second is created more slowly. Hence, CEP52 is inclined towards disulphide bridge formation.

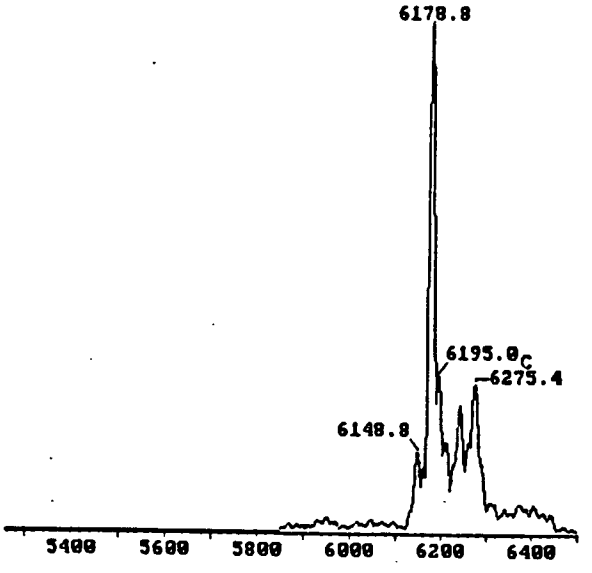


Figure 2.50 Electrospray Mass Spectrometry on CEP52

2.6.4 Purification and Characterization of UbCEP52

Purification of the 128 amino acid UbCEP52 was always going to present a greater challenge than the relatively small CEP52. The scale of the problem became evident during initial cleavage studies (see figure 2.51). Although the optimum cleavage conditions were identical to those used with CEP52, 6 hours under N_2 in aqueous TFA with EDT, EMS, anisole and phenol scavengers, the resulting broad HPLC peak revealed the high level of contamination in the crude material. As in the purification of CEP52, the five cysteine residues were initially left protected. Crude UbCEP52 was dissolved in a buffer containing 8 M urea and 50 mM NH_4OAc at pH 6.0 before being applied to a Sephadex G75 gel filtration column pre-equilibrated with the same buffer. Following elution from this column the urea was slowly dialysed from the protein solution. A molecular weight cut off of 10000 Da ensured the removal of both urea and a great many truncated peptide impurities. After dialysis the material was freeze dried and the five cysteines deprotected by treatment with aqueous tri-butylphosphine (3 hours under N_2). Fully deprotected

UbCEP52 was then dissolved in a buffer containing 10 mM DTT and 50 mM NH_4OAc at pH 6.0 and subsequently subjected to a second round of gel filtration. Analysis of the isolated fraction by SDS PAGE revealed it to contain a heterogeneous mixture of polypeptides (figure 2.53, lane 2) and clearly further purification was necessary.

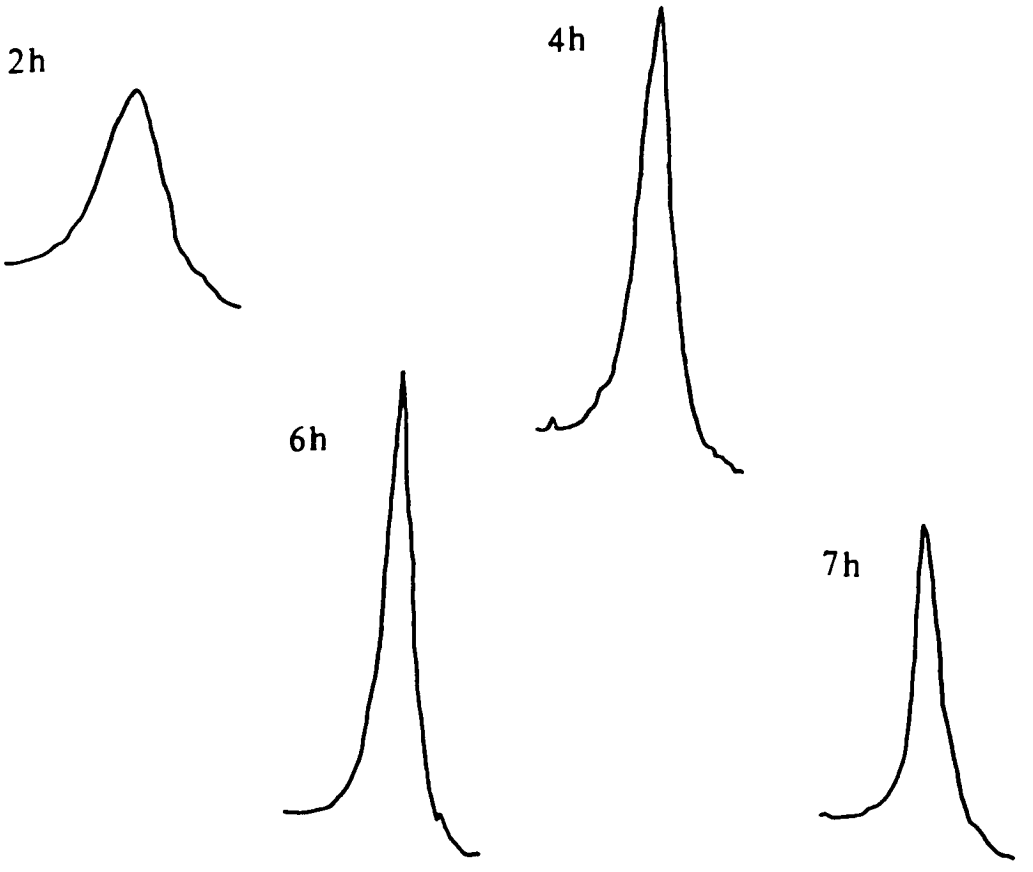


Figure 2.51 HPLC Traces obtained during Trial Cleavage of UbCEP52 C_8-RP300 column and gradient system A

Ion exchange chromatography seemed the obvious way of removing the impurities from the crude UbCEP52 mixture. Efficient use of this technique requires that the pI of the peptide be known and the best way of determining this is to use isoelectric focussing electrophoresis. UbCEP52 contains a high proportion of basic lysine (17) and arginine (10) residues and indeed carries an apparent nett charge of +14. It was therefore no surprise when the material focussed off the end of the IEF pH 3.5 - 9.5 gel, indicating that UbCEP52 has a pI greater than pH 9.5. In an early attempt at cation exchange chromatography, the synthetic protein was observed to bind

irreversibly to the CM-Sepharose matrix. This is in spite of the fact that the 50 mM to 1 M NH_4OAc salt gradient was carried out at pH 9.5. After this initial set-back, analytical cation exchange HPLC was used as a rapid means of searching out a more appropriate set of ion exchange conditions. However, despite screening from pH 4.0 to 11.0 no suitable conditions could be found. The protein would either remain stubbornly bound to the column or alternatively elute as a broad band. DNA binding proteins, themselves highly basic, are often purified using the weak cation exchange matrix heparin-agarose and it seemed worthwhile to try this out on UbCEP52. Unfortunately, as with many of the previous attempts, the synthetic protein could not be desorbed from the column, even when employing a 50 mM to 2 M $(NH_4)_2SO_4$ salt gradient at pH 8.0.

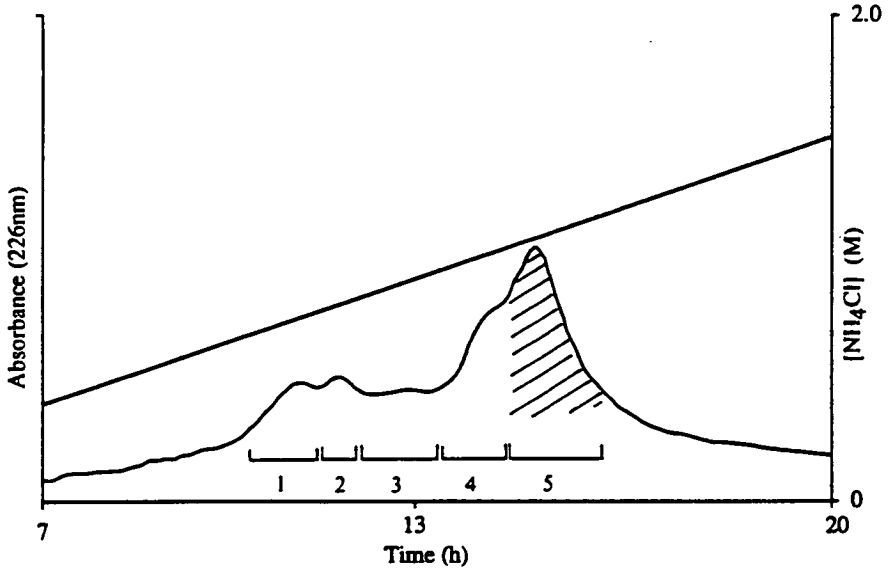


Figure 2.52 CM-Sepharose Cation Exchange on UbCEP52

In their work with the bovine 80 amino acid extension protein, Redman and Rechsteiner had found that buffers containing formic acid could be successfully used with cation exchange to remove contaminants from this naturally isolated protein [101]. Accordingly, UbCEP52 was loaded onto a CM-Sepharose column pre-equilibrated with a buffer containing 50 mM NH_4Cl, 1 mM DTT, 0.1 mM $ZnCl_2$ and 50 mM formic acid at pH 4.0. A subsequent salt gradient running from 50 mM to 1.5 M NH_4Cl was found to elute the various components of the mixture from the column (see figure 2.52). Note that a similar buffer containing 50 mM acetic acid at pH 4.0 had previously been tried, unsuccessfully, with analytical cation exchange

HPLC. At this time it is not known why the inclusion of the formic acid should make such a difference. The five fractions isolated from the CM-Sepharose column were analysed by homogeneous 20% SDS-PAGE which clearly revealed the extent to which the cation exchange chromatography had resolved the components of the crude mixture (see figure 2,53).

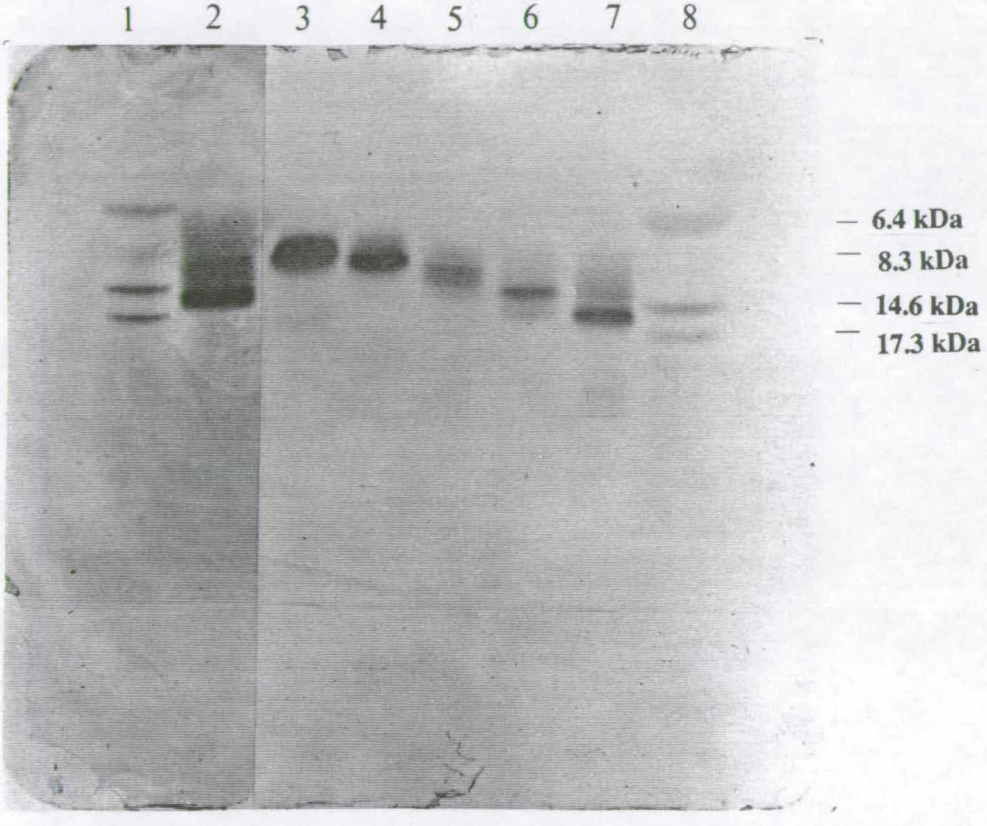


Figure 2.53 Homogeneous 20% SDS-PAGE Analysis of UbCEP52
lane1 & 8 - molecular weight markers
lane2 - crude fully deprotected UbCEP52
lane3 - CM-Sepharose F1
lane4 - CM-Sepharose F2
lane5 - CM-Sepharose F3
lane6 - CM-Sepharose F4
lane7 - CM-Sepharose F5

The major band in lane7 (F5) has an apparent molecular weight more or less consistent with that expected for UbCEP52 (theoretical mass 14728.59 Da).

Moreover, a simple inspection of the gel suggested this material was at least 80% homogeneous. Further purification would undoubtedly have removed some of the remaining contaminants, however, given the limited supply of F5 (16 mg) it was decided to accept this level of purity and so the purification was stopped at this point.

Acid hydrolysis of various duration were carried out on F5 and the resulting amino acid analyses are shown in table 2.14. By and large the amino acid composition of the synthetic material was as expected for UbCEP52. Again the cysteine residues appeared to oxidise to cysteic acid, but encouragingly the methionine content was not much lower than expected. This is of special significance given that the final residue to be coupled was a methionine.

Amino Acid	Expected	48 hour	72 hour	96 hour
Asx	12	11.2	11.88	12.47
Thr	8	7.31	6.51	6.12
Ser	4	5.66	4.60	4.27
Glx	15	18.90	18.40	18.20
Pro	6	-	6.36	7.74
Gly	7	-	-	-
Ala	5	5.17	6.83	6.44
Cys	5	3.42	2.38	1.57
Val	6	5.42	5.67	5.70
Met	2	1.70	1.50	1.54
Ile	10	8.94	8.83	8.79
Leu	13	12.62	12.30	12.70
Tyr	3	2.95	2.90	2.95
Phe	2	1.99	1.80	1.98
His	3	3.19	1.17	3.35
Lys	17	15.54	15.50	16.12
Arg	10	9.55	9.74	9.34

Table 2.14 Amino Acid Analysis of Synthetic UbCEP52
Times as indicated, 6N HCl (5 mg/ml Na_2SO_3) at 110°C

The primary sequence of the protein was further characterized by automated Edman degradation which confirmed the N-terminal residue was methionine [221]. Indeed, the first 30 residues in the protein were unambiguously sequenced and found to be in agreement with those in UbCEP52. Although circumstantial, there was evidence that the last 52 amino acids in UbCEP52 were also present and correct. The major product after the first part of the synthesis had already been shown to be CEP52 (section 2.6.3) and given that the construction of UbCEP52 was merely a continuation of this synthesis, it is fair to assume that the final product will also contain this 52 residue segment.

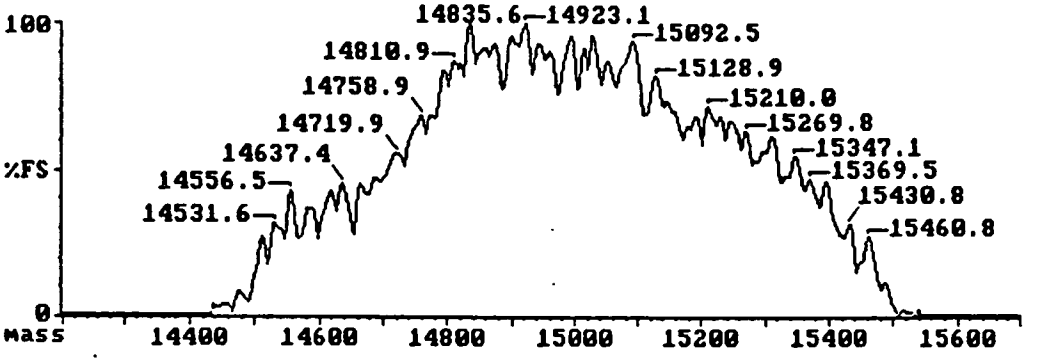


Figure 2.54 Electrospray Mass Spectrometry on Synthetic UbCEP52
Theoretical mass = 14728.59 Da

Electrospray mass spectrometry was performed on the synthetic material, however, the results it afforded could not be easily interpreted [195]. The technique is known to have some difficulty dealing with samples containing salts and this may account for the incongruous data it produced for UbCEP52 (see figure 2.54).

Laser desorption mass spectrometry gave a slightly more realistic mass determination for the synthetic material [224]. However, the experimentally obtained value of 14836 Da is still 107 mass units above that expected for UbCEP52 (see figure 2.55).

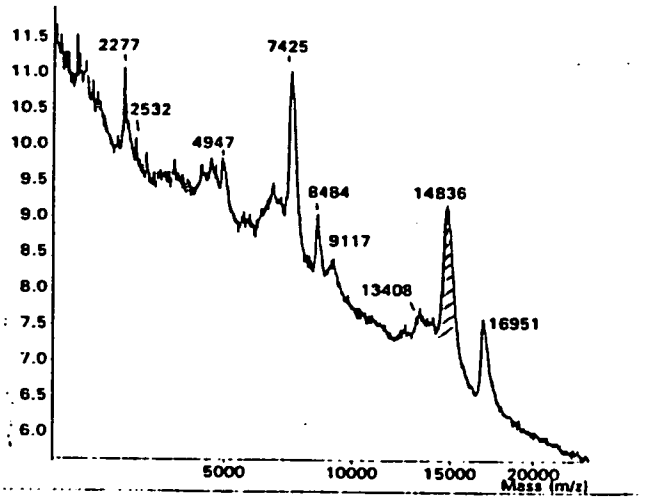


Figure 2.55 Laser Desorption Time-Of-Flight Mass Spectrometry on Synthetic UbCEP52
14836 & 7425 = M^+ & M^{2+} of synthetic UbCEP52
16951 & 8484 = M^+ & M^{2+} of apomyoglobin standard

During the final step of the purification the protein was exposed to $ZnCl_2$, thereby giving it the opportunity to bind the metal ion. Indeed, an Ellmans assay indicated the isolated protein contained 0.13 free thiol groups, suggesting either zinc binding or some other form of sulphydryl protection (eg disulphides). Zinc has an atomic weight of 65.38 Da, therefore a single coordinated ion could not itself account for the observed mass difference. If, however, the protein had been acetylated during the purification (mass change = +42 Da) then the combination of this and the metal coordination would explain the increase in mass. On the other hand, given the accuracy of the instrument in this mass range (± 10 Da), there are a number of other chemical modifications that could rationalise such a finding, for example, trifluoroacetylation, poly-formylation, various forms of alkylation etc. Alternatively, the mass determination itself may be at fault; the mass of the protein was calculated based on its time-of-flight relative to the internal apomyoglobin standard. Further mass studies using different internal standards are currently being carried out.

2.6.5 *In Vitro* Biological Activity of UbCEP52

Preliminary studies into the biological activity of UbCEP52 have been carried out by Mayer and co-workers [225]. All UbCEPs are rapidly processed *in vivo* by specific amidases, liberating both mature ubiquitin and the free CEP. This processing event served as the basis of the *in vitro* assay carried out on synthetic

UbCEP52. Presumably the amidase responsible for the cleavage would not recognize the processing site in UbCEP52 if the synthetic protein was in way defective in its primary or tertiary structure.

Anti-ubiquitin polyclonal antibodies (Dako Biochemicals) were found to recognize synthetic UbCEP52 (see figure 2.56, lane 4) thus allowing the putative processing of the precursor to be easily monitored by Western analysis. Human brain tissue was lysed in phosphate buffered saline (pH 7.0) and the resulting soluble protein fraction (S2) incubated with UbCEP52 for 2 hours at 37°C. The contents of the assay were then separated by SDS-PAGE and probed with anti-ubiquitin polyclonal antibodies (see figure 2.56).

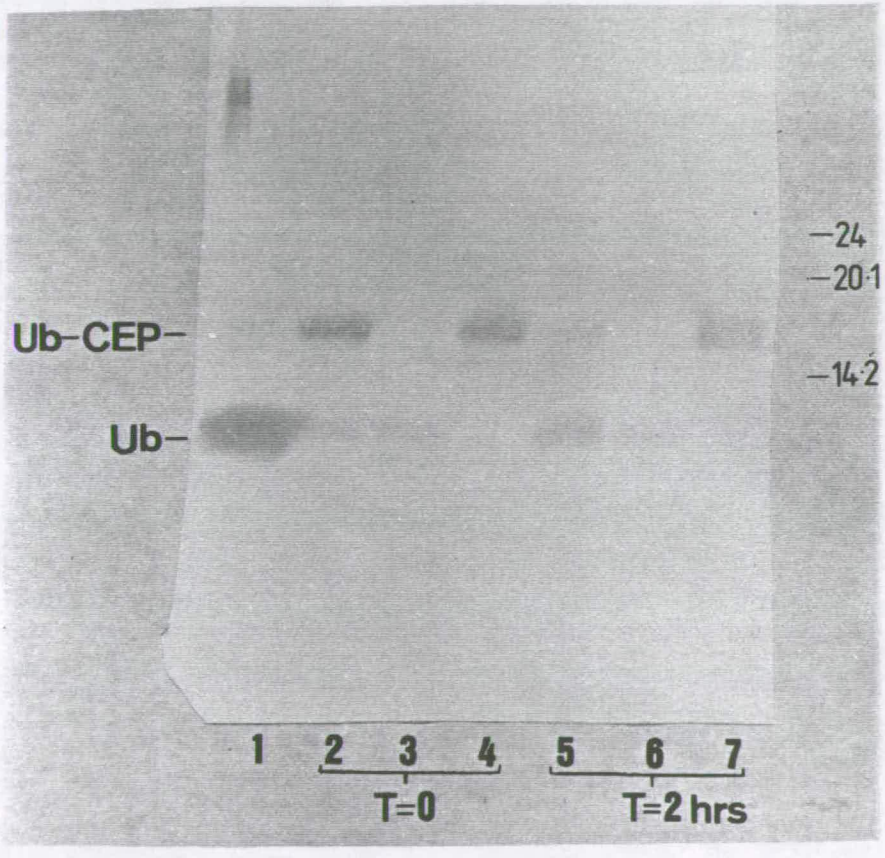


Figure 2.56 Western Analysis of the *In Vitro* Processing Assay on UbCEP52

lane1 = Bovine Ub
lane2 = UbCEP52 + S2 at T=0
lane3 = S2 at T=0
lane4 = UbCEP52 at T=0
lane5 = S2 + UbCEP52 at T=2 h
lane6 = S2 at T= 2 hrs
lane7 = UbCEP52 at T=2 h

Comparison of lanes 2 and 5 reveals that ubiquitin is liberated from the UbCEP52 after a 2 hour incubation with human brain soluble protein fraction. Synthetic UbCEP52 was also shown to be processed when incubated with rabbit reticulocyte lysate, implying that a similar amidase is present in both human and murine tissues. Naturally this assay is not able to detect the liberated CEP52 since this peptide is not immunoreactive against anti-ubiquitin antibodies. Therefore the possibility that CEP52 is digested by protease in the system cannot be completely ruled out. However, by radiolabelling the 3 tyrosines in UbCEP52 with ^{125}I, both products of the processing can be detected by autoradiography. Furthermore, the extreme sensitivity of autoradiography should make it possible to detect any endogenous protein conjugates the mature ubiquitin might form after the processing event. The presence of such conjugates would of course be strong evidence that the ubiquitin component had folded into its native conformation. Studies involving radiolabelled UbCEP52 are currently in progress [225].

2.6.6 Structural Studies on CEP52 and UbCEP52

The principal aim of this study was to investigate both the zinc binding and structural properties of CEP52 and UbCEP52. To do so we employed a combination of CD spectroscopy, UV spectroscopy and NMR spectroscopy. Where appropriate the results were compared with those obtained for purified bovine ubiquitin.

2.6.6(i) Circular Dichroism Studies on CEP52 and UbCEP52

The far UV CD spectra of CEP52 and UbCEP52 were acquired in various buffered solutions and in each case the secondary structure content determined using the CONTIN procedure (see table 2.15). In aqueous boric acid at pH 6.0, neither CEP52 nor UbCEP52 had a CD spectrum consistent with a high amount of secondary structure. Indeed, it was significant that UbCEP52 apparently had less α-helix than mature ubiquitin (see figure 2.57). No obvious change was observed in the CD spectra of CEP52 upon the addition of $ZnCl_2$ (10 eq). Zinc finger peptides commonly contain between 20 and 40% α-helix, therefore CEP52 did not appear to be adopting a structure analogues to these peptides. The possibility that disulphide bridge formation was preventing zinc binding was investigated by adding dithiothreitol, however, inclusion of this reducing agent had no conspicuous effect on the secondary structure of the peptide.

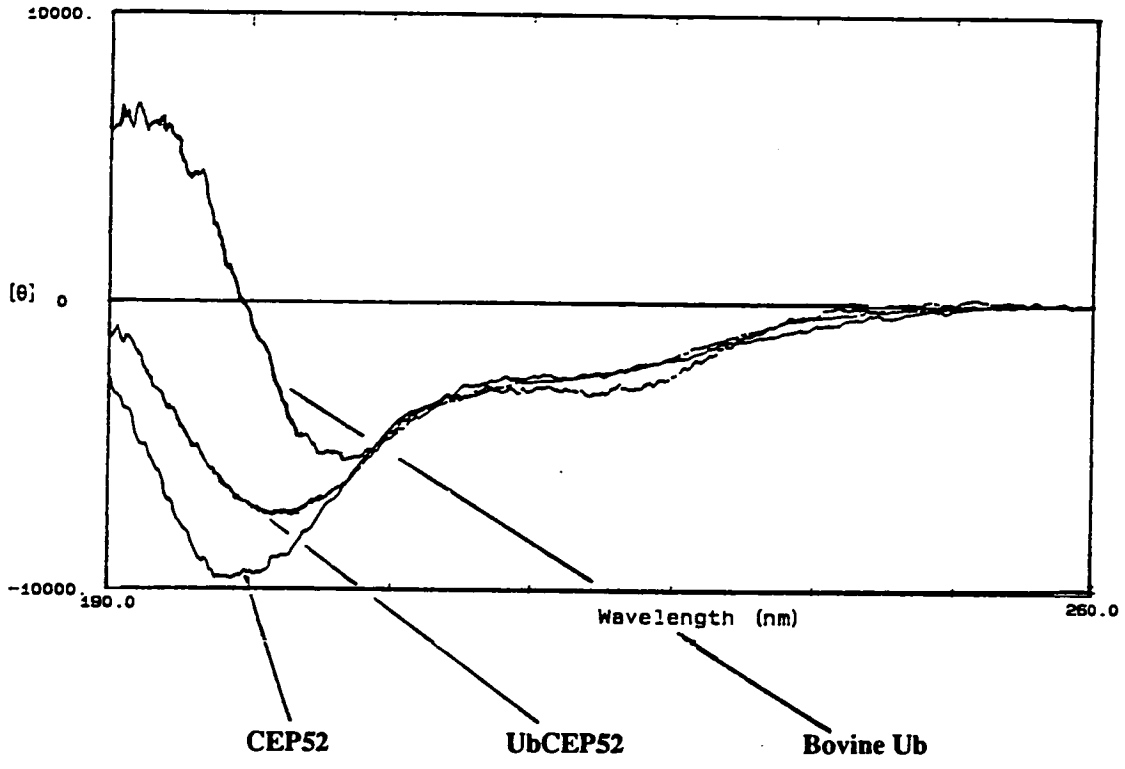


Figure 2.57 Far UV CD Spectra of Bovine Ubiquitin, CEP52 and UbCEP52 in 50 mM Boric Acid at pH 6.0

UbCEP52 had already been exposed to zinc during the purification of the protein, thus making it impossible to study the effect the metal ion had on its structure. Nevertheless, the α–helix content in UbCEP52 exposed to zinc was still substantially lower than that in a classical zinc finger motif. Again, addition of DTT had no effect on the level of helicity in the protein.

Exposure of the samples to trifluoroethanol (50% by volume) dramatically altered their far UV CD spectra. In each case an appreciable increase in α-helicity was associated with the addition of this alcohol (see table 2.15). Indeed, both CEP52 and UbCEP52 exhibited CD spectra more akin to those expected for zinc finger type motifs. However, such an artificial environment also induced a considerable increase in the α-helicity in bovine ubiquitin, raising questions over the ultimate importance of these TFE generated structures.

Sample	α-helix	β-sheet	Remainder
Bovine Ub	15 ± 2.5%	57 ± 2.6%	28 ± 4.5%
+ 50% TFE	63 ± 0.73%	38 ± 0.73%	0 ± 7.1 x 10^{-7}
CEP52	3 ± 0.48%	38 ± 0.55%	59 ± 0.95%
+ Zn^{2+}(10eq) + DTT (5eq)	4 ± 0.80%	42 ± 0.83%	54 ± 1.5%
+ Zn^{2+}(10eq) + DTT(5eq) +TFE	24 ± 0.37%	28 ± 0.54%	48 ± 0.66%
UbCEP52 + Zn^{2+}	3 ± 0.69%	55 ± 0.72%	42 ± 1.3%
+ 50% TFE	31 ± 1.2%	43 ± 1.2%	27 ± 2.2%

Table 2.15 Secondary Structure Contents Derived from CD using CONTIN Procedure: all samples dissolved in 50 mM boric acid at pH 6.0 containing additives stated

2.6.6(ii) Ultra-Violet Absorption Studies on CEP52 and UbCEP52

Coordinated transition metal ions very often give rise to distinctive absorption bands in the UV-visible region of the electromagnetic spectrum. These bands are produced in two distinct ways. The first, d-d transitions, involve discrete electronic shifts between the split d-orbitals of the coordinated metal centre. All d-d transitions are forbidden processes ($\Delta l=0$) and for this reason are generally weak in intensity. On the other hand, charge-transfer bands, which involve transitions between the molecular orbitals of the ligand and the metal d-orbitals, are allowed transitions and are thus of greater intensity. Tetrahedrally coordinated Zn^{2+} (d^{10}) does not have any d-d transitions and therefore no identifying absorption bands in the UV-visible region. Fortunately tetrahedrally coordinated Co^{2+} (d^7) does have a characteristic d-d transition in the 600-700 nm region, and for this reason is often used in place of Zn^{2+} in UV-visible binding studies [226]. Furthermore, if the Co^{2+} is coordinated to cysteinate ligands there will also be strong charge-transfer bands in the 300-350 nm region of the spectrum [226].

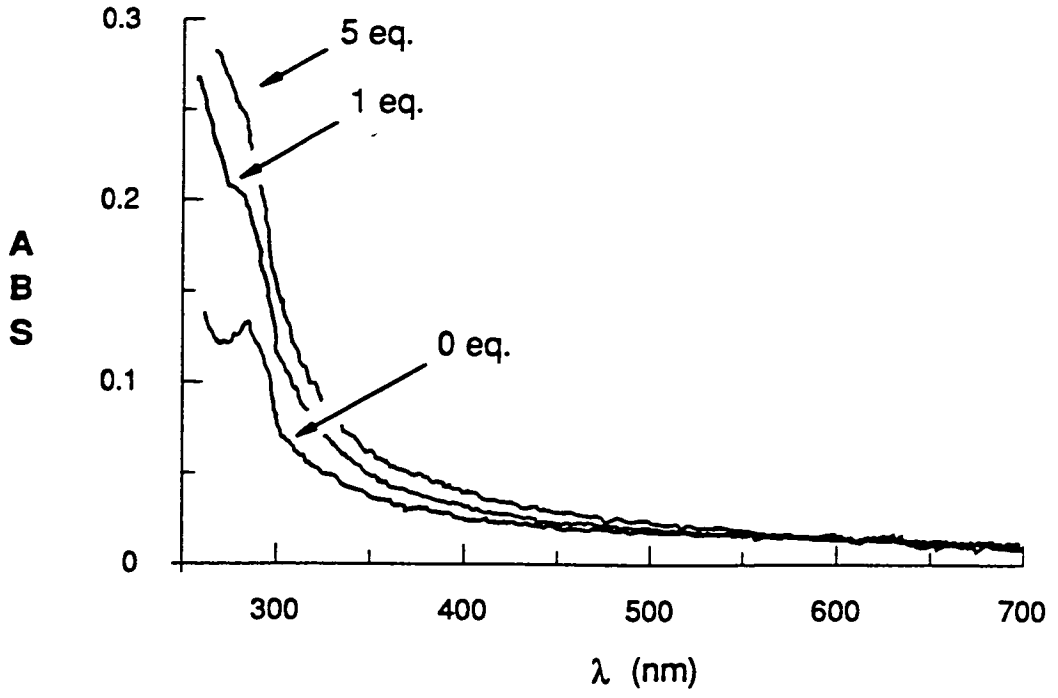


Figure 2.58 **Optical Absorption Spectra of Reduced CEP52 with and without Co^{2+}: (a) 50 μM CEP52; (b) 50 μM CEP52 + 1eq $CoCl_2$; (c) 50 μM CEP52 + 5eq $CoCl_2$**
Spectra (b) and (c) have been corrected by subtracting the spectrum of free $CoCl_2$

Samples of CEP52 and UbCEP52 were sonicated overnight at 40°C in a buffer containing 50 mM Tris.HCl at pH 7.2, 10 mM DTT and 10 mM EDTA. The reduced samples were then recovered by semi-preparative HPLC and immediately lyophilised ready for use. UV-visible analysis significantly revealed that addition of $CoCl_2$ to CEP52 was not accompanied by the appearance of the d-d band in the required 600-700 nm region (see figure 2.58). Similarly, no such band was observed in the UV-visible spectrum of UbCEP52 upon addition of $CoCl_2$ (see figure 2.59). These results appear to indicate Co^{2+} is not tetrahedrally complexed in either CEP52 or UbCEP52, thereby implying the two samples do not bind Zn^{2+} in a tetrahedral manner.

Interestingly, both CEP52 and UbCEP52 have a noticeable increase in absorbance in the 300-350 nm region upon addition of the $CoCl_2$. Such an absorbance appears to be consistent with the cysteinate-Co^{2+} charge-transfer band usually associated with zinc and cobalt binding peptides. Why this charge-transfer band should be

present and not the d-d transition is difficult to explain, although it is possible that the d-d transition is so weak as to be virtually invisible.

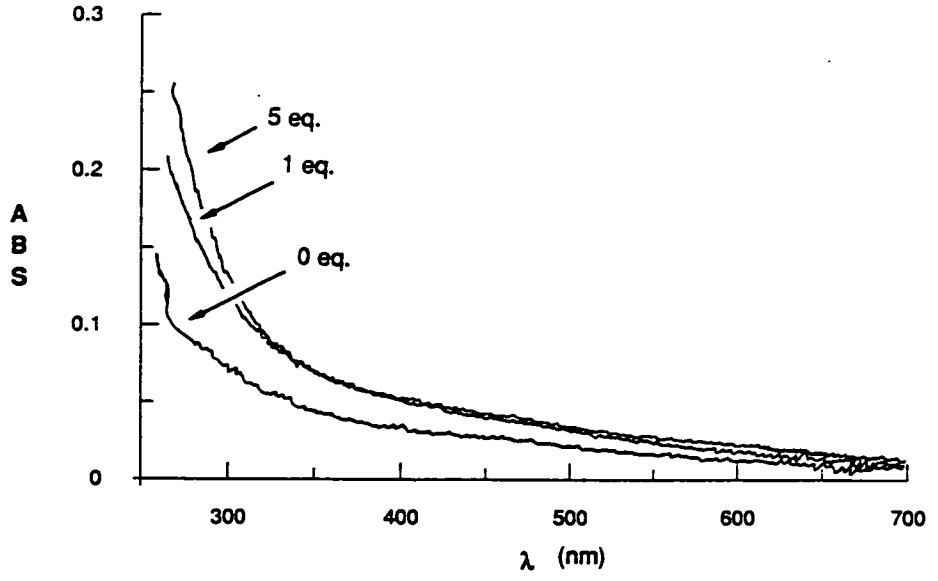


Figure 2.59 Optical Absorption Spectra of Reduced UbCEP52 with and without Co^{2+}: (a) 25 μM UbCEP52; (b) 25 μM UbCEP52 + 1eq $CoCl_2$; (c) 25 μM UbCEP52 + 5eq $CoCl_2$
Spectra (b) and (c) have been corrected by subtracting the spectrum of free $CoCl_2$

2.6.6(iii) NMR Studies on CEP52 and UbCEP52

The sensitivity of NMR to changes in peptide/protein structure makes it ideally suited to monitoring the effect an additive, like $ZnCl_2$, has on the conformation of such a molecule. *In situ* titration of a sample of reduced CEP52 with $ZnCl_2$ was followed by 1D NMR, the results of which are shown in figure 2.60. Addition of the metal ion did not appear to bring about any great change in the amide proton region, although there were some minor differences in the 8.4 - 8.6 ppm range of the spectrum. Note, the 1D ^1H NMR spectrum of CEP52 + $ZnCl_2$ (10eq) + DTT (5eq) was no different from that of CEP52 + $ZnCl_2$ (1eq). The tetrahedral coordination of Zn^{2+} in peptides normally results in considerable conformation reorganization, and hence conspicuous changes in the 1D NMR spectrum of the peptide [225]. Therefore, the small differences observed in the 1D NMR spectrum of CEP52 must be regarded as insignificant in comparison with those usually associated with Zn^{2+} binding. Consequently, there does not seem to be any

evidence from NMR to support the contention that CEP52 is a zinc binding peptide. Indeed, this conclusion is entirely in keeping with the observations from circular dichroism and ultra-violet absorption spectroscopies.

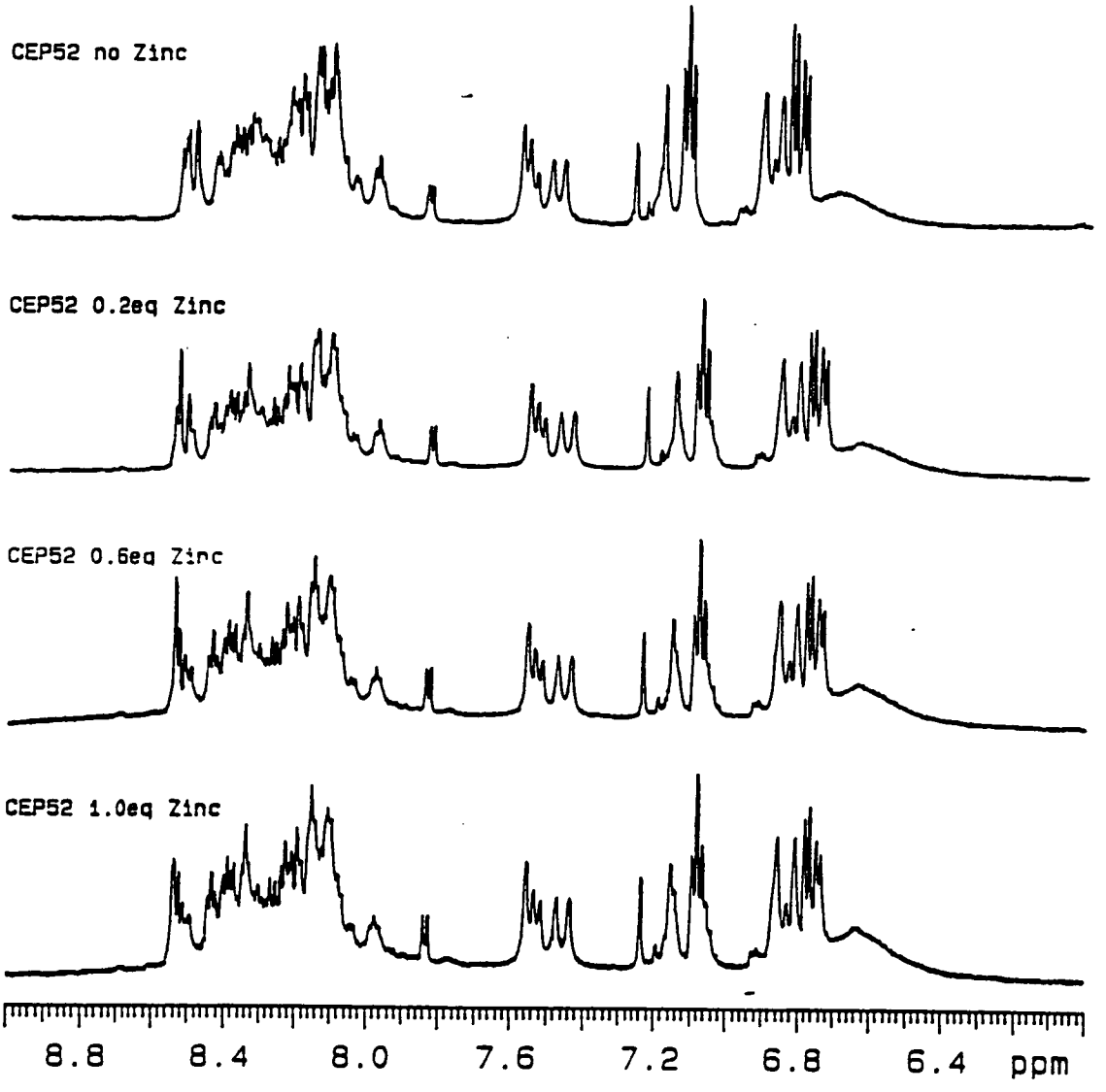


Figure 2.60 **1D ^{1}H NMR Spectra of CEP52 indicating the effect of $ZnCl_2$ addition. Spectra recorded at 25^{o}C in 90% H_2O: 10% D_2O at pH 6.0**

Two dimensional NMR was used to study in more detail the solution conformation of CEP52 and UbCEP52. The complete lack of NH - αCH crosspeak dispersion in the TOCSY spectrum of CEP52 in the presence of $ZnCl_2$ (10eq) confirmed the absence of well defined structure earlier suggested by CD (see figure 2.61).

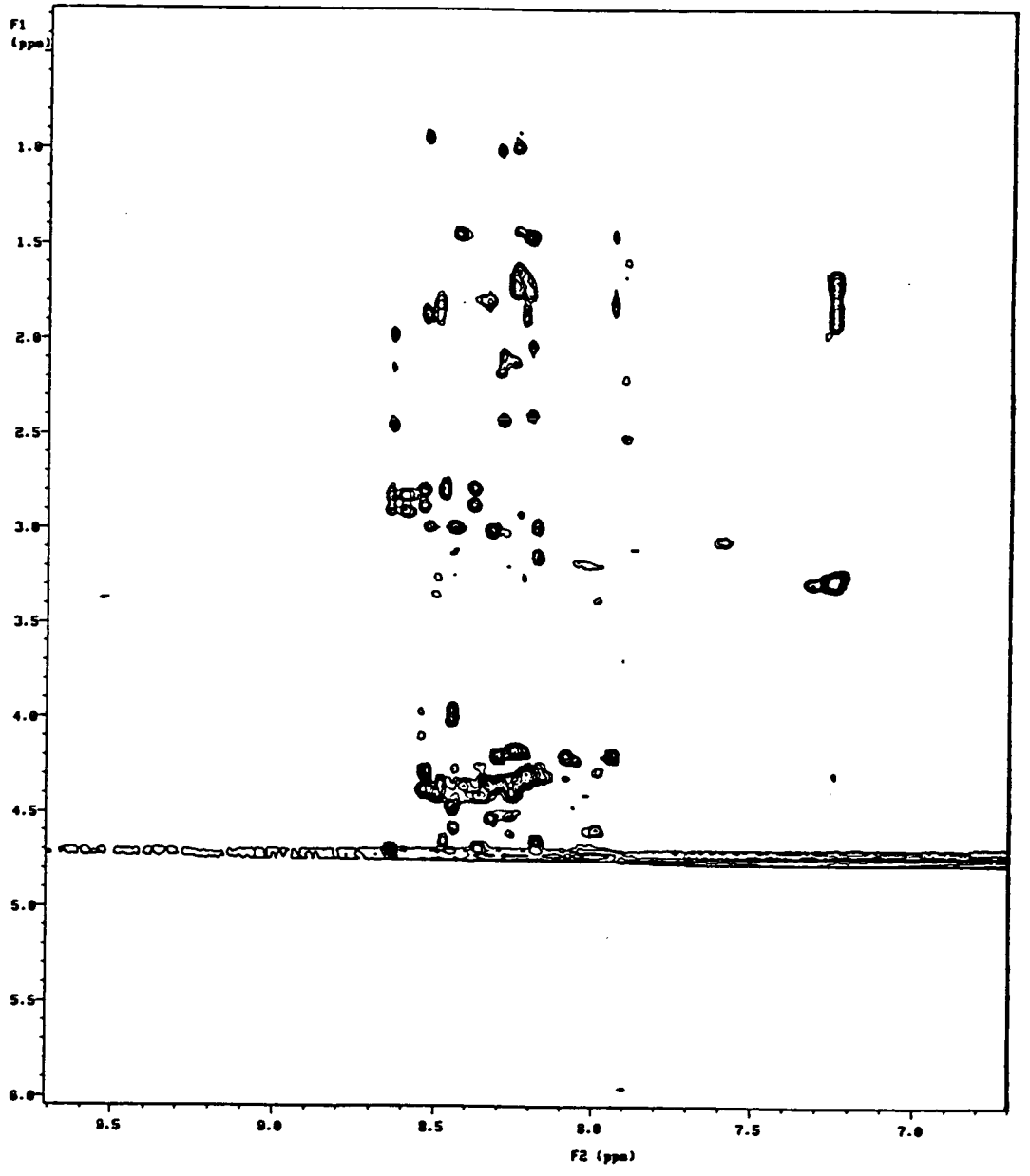


Figure 2.61 Region from the TOCSY Spectrum of CEP52 in 90% H_2O: 10% D_2O at pH 6.0 containing 10eq $ZnCl_2$, 25°C

Interestingly, the TOCSY spectrum of CEP52 did not appear to change when the peptide was exposed to 50% TFE, an observation apparently in disagreement with the previous CD studies. The possibility that in this medium the peptide exhibits isolated areas of secondary structure, more easily detected by CD than NMR, cannot be disregarded.

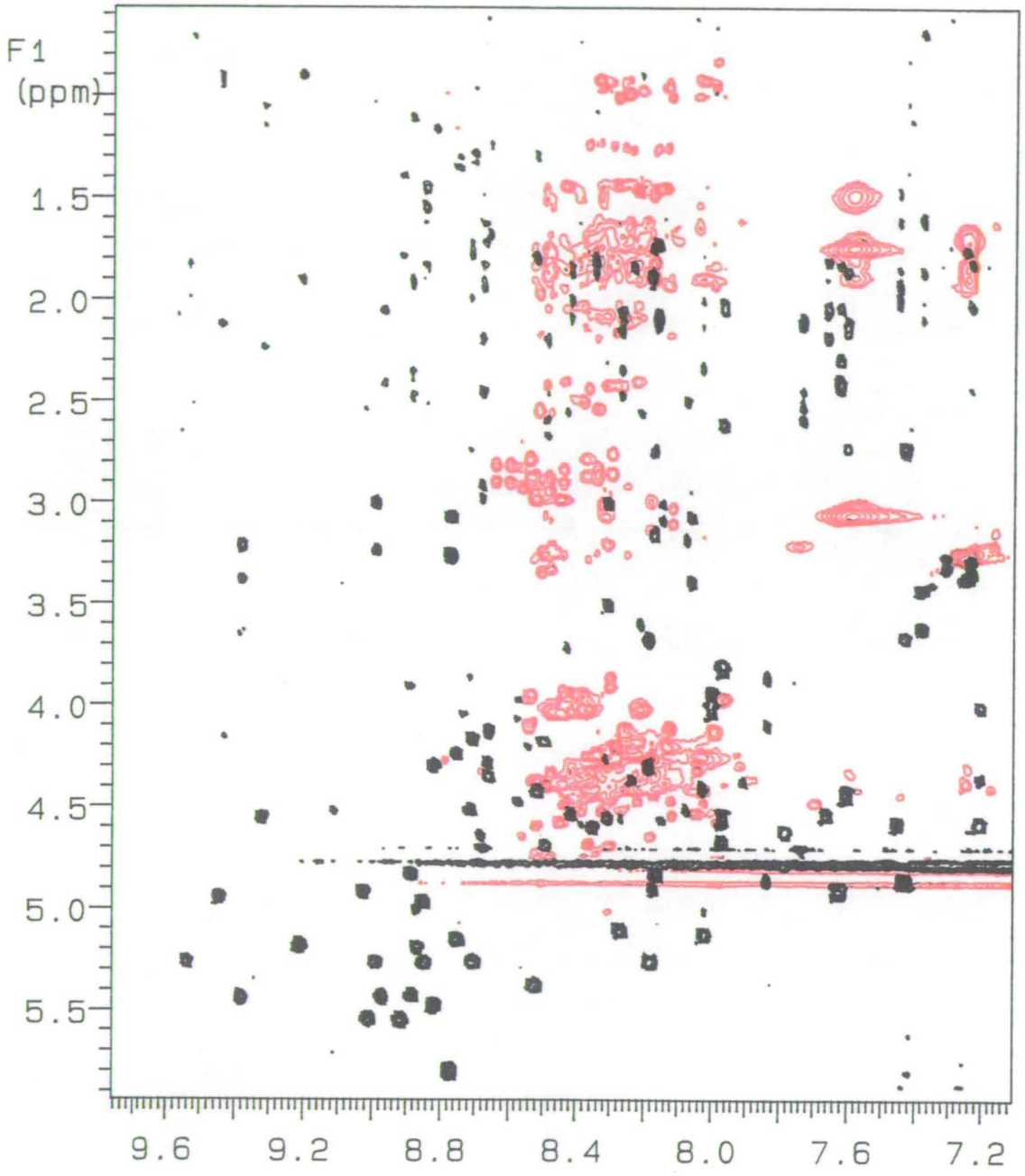


Figure 2.62 Region from the TOCSY Spectra of UbCEP52 (red) and Bovine Ubiquitin (black). Spectra obtained in 90% H_2O: 10% D_20 at pH 6.0 containing 10eq $ZnCl_2$, 25°C

Although the role of the ubiquitin component in UbCEP52 is not absolutely certain, it is not unreasonably to suggest that it in some way allows the CEP to adopt its native conformation. Indeed, given that CEP52 appeared to be essentially random coil in aqueous solution, this function did have some attraction. It therefore came as some surprise when the TOCSY spectrum of UbCEP52 not only indicated the CEP52 component had no structure, but also the ubiquitin segment had unfolded (see figure 2.62). As in the case of CEP52 exposure of the protein to 50% aqueous TFE had no visible effect on the TOCSY spectrum.

The random coil nature of CEP52 and UbCEP52 indicated by the exceptional spectral crowding and overlap, made assignment of their proton resonances impossible.

2.6.7 Conclusions

All the evidence accumulated from these structural studies seems to suggest that neither CEP52 nor UbCEP52 are able to bind Zn^{2+} in a tetrahedral manner. *In situ* disulphide formation would of course prevent these peptides from coordinating Zn^{2+}, and hence explain their apparent indifference to the addition of this metal ion. However, the random coil nature of CEP52 and UbCEP52, as indicated by both CD and NMR, would seem to rule out the existence of structure inducing disulphide bridges. Initial biochemical investigations indicate that UbCEP52 is correctly processed into mature ubiquitin and CEP52 *in vitro*. Presumably the amidase responsible for this process is sensitive to the structural surrounding the cleavage site, and hence is happy with the random coil nature of synthetic UbCEP52. It is not inconceivable that ubiquitin unfolds within UbCEPs so as to allow this amidase easier access to the processing site. This unfolding of ubiquitin within UbCEPs would surely be a handicap to the folding of the CEP component.

Whether or not this putative ubiquitin induced precursor unfolding is a general phenomenon remains to be seen. However, given the relatively short *in vivo* half-lives of many heterologous gene products encoding ubiquitin N-terminally fused to various proteins [106-108], it is easy to speculate that being tethered to ubiquitin prevents these proteins from folding and so facilitates rapid *in vivo* processing of such precursors.

Chapter 3 - Experimental

Section 3.1 Notes

All Fmoc protected amino acids were purchased from Novabiochem and Raylo (Fmoc-Arg(Pmc)OH) and unless otherwise stated were of the L-configuration. ^{15}N-Glycine was purchased from Sigma and used as supplied. Melting points were recorded in open capilliaries on a Buchi 510 melting point apparatus and are uncorrected. Thin layer chromatography (t.l.c.) was carried out on commercially available plastic sheets pre-coated with silica gel 60 F_{254} (Merck) using the following system: $CHCl_3$ / MeOH / AcOH (9:1:0.5). Visualization of the compounds was achieved by ultra-violet absorption at 254 nm and ninhydrin in the case of free amino groups. Infrared spectra were recorded on a Perkin Elmer 781 double beam spectrophotometer in the solvent indicated. Polystyrene (1603 cm^{-1}) was used as the standard. Ultra-violet spectra were recorded on a Varian Cary 210 spectrophotometer in the solvent indicated. Where necessary temperature control was achieved by continuously pumping water from a thermostatic bath (Techne TE-8J) through the thermostated cell compartment of the instrument. Temperatures were measured using a Comark electronic thermometer. Refractive indices were measured on an Abbe 60 refractometer. A Corning ion analyser 150 was used for all pH measurements. Circular dichroism (CD) spectra were recorded on a JASCO J600 spectropolarimeter in the solvent indicated. Fluorescence spectra were recorded on a Perkin Elmer LS50 luminescence spectrometer fitted with a Grant LTD-6 thermostatic control. Proton NMR spectra were acquired on a Varian VXR 5000 (600 MHz) instrument in the solvent indicated and the chemical shifts were measured relative to TMS assigned to zero. High and low resolution fast atom bombardment (FAB) were measured on a Kratos MS5OTC machine. High performance liquid chromatography (HPLC) was carried out using either a Waters/Kratos system: 2 x Kratos spectroflow 400 pumps, a Waters U6K injector, a Waters 680 automatic gradient controller and a Waters 450 variable wavelength detector; or one of two Applied Biosystems systems: A) 2 x 400 solvent delivery systems, a 491 injector/mixer and a 783A detector/programmer, B) 2 x 1406A solvent delivery systems, a 1480A injector/mixer and a 1783A detector/programmer. Analytical and semi-preparative separations were performed using the conditions and columns described in the text. Amino acid analysis was carried out on an LKB 4151 amino acid analyser following either sealed Carius tube hydrolysis with 6N HCl (containing 5 mg/ml Na_2SO_3) at 110^{o}C for the time indicated in the text, or total enzymatic hydrolysis using the enzymes indicated in

the text. Protein sequencing was performed by the Welmet protein characterization group on an Applied Biosystems 477A protein sequencer. Fast protein liquid chromatography (FPLC) was carried out on a Pharmacia instrument using the conditions and columns indicated. Isoelectric focussing electrophoresis (IEF) was performed on a LKB 217 multiphor horizontal bed instrument powered by a Pharmacia ECPS 3000/150 constant powerpack using the conditions and gels stated in the text. SDS-PAGE was carried out on a Pharmacia phastsystem using precast 20% homogeneous SDS phastgels and buffer strips. All gels were stained with Coomassie blue R250. Gel filtration and ion exchange chromatography were carried out using Pharmacia/LKB apparatus: 2 x LKB 2138 UVICORDS, a Pharmacia 2132 Microperpex peristaltic pump, an LKB 2112 redirac fraction collector and where appropriate a Pharmacia GM gradient mixer. All separations employed the columns and conditions stated in the text. All solvents were distilled before use and the following were dried using the reagents given in parenthesis when required: dichloromethane (calcium hydride), diethyl ether (sodium wire), ethanol (magnesium iodine). The dimethylformamide (DMF) and 1,4-dioxan used were peptide synthesis grade, supplied by Rathburn Chemicals.

Section 3.2 Solid Phase Peptide Synthesis

All peptides described were synthesised on an Applied Biosystems 430A automated peptide synthesiser equiped with an Applied Biosystems 757 absorbance detector linked to a Hewlett Packard HP3396A Integrator for online monitoring of the deprotection solution containing the piperidine-fulvene adduct.

An orthogonal synthetic strategy was employed in which the base-labile 9-fluorenylmethoxycarbonyl (Fmoc) group was used for N^{α}-amino protection. This allowed both acid-labile side-chain protection and acid-labile peptide-resin linkage. The following side-chain protecting groups were used: tertiary-butyl (Bu^t) ethers for serine, threonine and tyrosine; Bu^t esters for glutamic and aspartic acid; 4,4'-dimethoxybenzhydryl (Mbh) for asparagine and glutamine; tertiary-butyloxycarbonyl (Boc) for lysine; τ-triphenylmethyl (Trityl, Trt) protection for histidine and 2,2,5,7,8-pentamethylchroman-6-sulphonyl (Pmc) protection for the guanidino group of arginine. The sulphydryl group of cysteine was protected with the non-acid-labile tertiary-butylthio (S-Bu^t) group.

The first (C-terminal) amino acid was attached to the acid sensitive *p*-alkoxybenzyl alcohol (Wang) group on the resin outwith the synthesiser. The syntheses were monitored using real-time assessment of the deprotection solution (2 ml aliquot) containing the chromophoric piperidine-fulvene adduct at 302 nm in a continuous flow mode from the synthesiser to the UV detector. Most amino acids were triple coupled, firstly as a symmetrical anhydride (2 mmol amino acid) followed by two 1-hydroxybenzotriazole (HOBt) active ester (1 mmol amino acid) couplings using N,N-diisopropylcarbodiimide (DIC) in dioxan as the activating agent. Exceptions to this were glycine which was single coupled as a symmetrical anhydride (4 mmol amino acid) and histidine, asparagine and glutamine which were coupled three times as their respective HOBt active esters (1 mmol amino acid). Following coupling any unreacted N^{α}-amino groups were capped by treatment with acetic anhydride.

Recurrent deprotection, activation, coupling and capping steps were carried out using the following programmed cycles.

1. **Deprotection:**
The Fmoc group was removed from the peptide-resin with successive 5, 3, 3 and 1 minute treatments with 20% piperidine in DMF. The resin was drained and washed twice with DMF between consecutive deprotections.

2. **Wash:**
The resin was washed 5 times with DMF.

3. **Coupling:**
(a) The protected amino acid was preactivated as a symmetrical anhydride using DIC in dioxan 15 minutes prior to being added to the resin. Coupling was then allowed to proceed in a 1:1 mixture of DMF:dioxan for 30 minutes.
(b) A second coupling of the amino acid pre-activated as an HOBt active ester for 20 minutes was carried out in a 1:1 mixture of DMF:dioxan for 30 minutes.
(c) The amino acid was once again coupled as an HOBt acive ester for 30 minutes in a 1:1 mixture of DMF and dioxan.

4. **Wash:**
The resin was washed 4 times with DMF.

5. **Capping:**

Capping of the resin was achieved with a mixture containing 1 ml of 0.5 M acetic anhydride in DMF and 1 ml of 0.5 M pyridine in DMF for 2 minutes, drained and washed twice then capped a second time for 4 minutes.

6. **Wash:**

The resin was washed 5 times with DMF.

Note, 1-hydroxybenzotriazole which is purchased (Fluka) as the hydrate (12-17%) was dried over anhydrous $MgSO_4$ and recrystallized from dry ethanol before being used. DIC and acetic anhydride were purchased from Aldrich and used as supplied.

Section 3.3 Experimental

Coupling the C-Terminal Amino Acid onto the *p*-Alkoxybenzyl Alcohol Resin

The first amino acid was attached to the polymeric support using one of the following manual procedures.

1. **Acid Chloride Coupling**

The Fmoc-amino acid (3.16 mmol) was suspended in dry DCM (50 ml) and thionyl chloride (2.20 ml, 30.16 mmol) added under nitrogen. The mixture was heated under reflux at 50°C for 2 hours after which the DCM and excess thionyl chloride were removed under vacuum. The resulting solid was then redissolved in DCM and concentrated *in vacuo*. This was repeated to give a pale yellow solid; ν_{max} 1710-1720 (COCl), 1800 (Fmoc C=O) cm^{-1}. *p*-Alkoxybenzyl alcohol resin (2.0g, 1.58 mmol) was pre-swollen in dry DCM (25 ml) and pyridine (3 ml) for 30 minutes before adding the Fmoc amino acid chloride (from above) dissolved in dry DCM (10 ml). After sonicating for 1 hour the resin was filtered, washed with DCM (3 x 50 ml), diethyl ether (3 x 50 ml) and dried in a dessicator overnight. The resin substitution and percentage loading was then determined.

2. **Diisopropylcarbodiimide Coupling**

The Fmoc-amino acid (1.2 mmol) was activated as the symmetrical anhydride by treatment with diisopropylcarbodiimide (DIC) (94 µl, 0.6 mmol) in DMF (10 ml). This mixture was sonicated for 15 minutes before being added to *p*-alkoxybenzyl alcohol resin (1.5 g, 1.2 mmol) previously swollen in DMF (20 ml). 4-Dimethylaminopyridine (DMAP) (10 mg, 0.09 mmol) was added and the resin

sonicated for 15 minutes. The resin was then filtered, washed with DMF, DCM and diethyl ether and dried overnight in a dessicator. The resin substitution was then determined.

Coupling method 1 was used only with Fmoc-glycine, all other amino acids were attached to the resin using the second procedure. Note that the reaction times and molar ratios of reagents in both coupling methods were chosen to give low initial loading on the resin.

Determination of Resin Loading

The loading efficiency was determined by sonicating a sample of the dried resin (3-4 mg) for 10-20 minutes in the presence of a 20% (v:v) piperidine in DMF solution (10 ml). The UV absorbance of the solution at 302 nm was then measured and the concentration of the piperidine-fulvene adduct therein determined using the Beer-Lambert law ($\varepsilon_{adduct,302\ nm}$ = 15400 $dm^3\ mol^{-1}\ cm^{-1}$). This value then allows the initial loading of the resin to be calculated.

Capping of Fmoc-Amino Acid-Resin

Fmoc-amino acid-resin (1.60 g) was suspended in DCM (30 ml) and cooled to 0°C. Pyridine (0.5 ml) and benzoylchloride (0.5 ml) were then added and the mixture stirred for 5 hours at room temperature. The resin was filtered, washed with DCM (3 x 50 ml), diethyl ether (3 x 50 ml) and dried in a dessicator overnight.

Glycylglycine Content in Loaded Resin

In the case where the first amino acid is glycine, it is necessary to check for unwanted polymerisation of this residue. A sample of the resin (20 mg) was treated with 20% piperidine in DMF (30 ml) for 20 minutes to remove the Fmoc group. Glycine was removed from the resin by treatment with 95% aqueous TFA for 1 hour. After removal of the TFA *in vacuo* the residue was triturated with diethyl ether to give a white powder which was isolated by filtration and dissolved in citric acid buffer at pH 2.2 for amino acid analysis. The retention time (R_T) of the amino acid in the sample was then compared with those of glycine (R_T = 6.19 min) and glycylglycine (R_T = 39.2 min) standards.

Determination of the Fmoc Content in Completed Peptide

The overall yield of the synthesis on the resin was determined in the following way. The final (N-terminal) amino acid was left with the Fmoc protecting group attached so as to allow UV monitoring. A sample of dried Fmoc-peptide-resin (2-5 mg) was sonicated for 10-20 minutes in the presence of 20% piperidine in DMF (10 ml). The UV absorbance of the solution at 302 nm was then measured and the concentration of the piperidine-fulvene adduct therein determined using the Beer-Lambert law. This value allows the final Fmoc-amino acid loading to be calculated and this can be compared with the theoretical final loading given 100% yield in every step of the synthesis (determined using computer programme).

Preliminary Acidolytic Cleavage of Resin Bound Peptide

Peptide cleavage conditions were optimized in the following way. Fmoc-peptide-resin was deprotected with 20% piperidine in DMF for 10-20 minutes before being washed with DMF, DCM and diethyl ether and dried overnight in a dessicator. Deprotected peptide-resin (30-50 mg) was swollen in a mixture of scavengers (0.25 ml, 5% v:v of each) for 30 minutes under nitrogen. Depending upon the peptide these scavengers would be chosen from the following: anisole; thioanisole; phenol; ethyl methyl sulphide (EMS); ethane dithiol (EDT). Trifluoroacetic acid (TFA) (5 ml, 95% v:v) and H_2O (0.25 ml, 5% v:v) were then added to the pre-swollen resin which was then gently stirred under nitrogen. Aliquots (0.75 ml) of the cleavage mixture were periodically removed and concentrated *in vacuo*. The resulting residues were triturated with diethyl ether to give white peptide precipitates. These were filtered off, dissolved in 20% aqueous acetic acid and analysised by HPLC.

High Performance Liquid Chromatography (HPLC)

Unless otherwise stated all analytical HPLC was performed on an Aquapore C_8 reverse phase column (100 x 4.6 mm, 300 A^o pore size, 7μm) employing a flow rate of 1 ml/min and monitoring at 214 nm. Semi-preparative HPLC was accomplished on an Aquapore C_8 reverse phase column (100 x 10 mm, 300 A^o pore size, 20 μm) using a flow rate of 5 ml/min and monitoring at 214 nm. The following gradient systems were used with both analytical ans semi-preparative HPLC.

Gradient System A:

A = H_2O, B = CH_3CN, 0.1% TFA, 10 - 90% B over 30 minutes.

Gradient System B:
A = H_2O, B = CH_3CN, 0.1% TFA, 20 - 40% B over 20 minutes.

Purification of Bovine Ubiquitin (Sigma)

Commercially available bovine ubiquitin was purified in the following manner.

1. **Gel Filtration Chromatography**

Bovine ubiquitin (93.7 mg, 10.9 μmol) was dissolved in a buffer containing 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5 (10 ml) and applied to a Sephadex G50 (fine) column (800 x 30 mm) pre-equilibrated with the same buffer. The column was then eluted with 8 M urea, 10 mM DTT and 50 mM NH_4OAc (pH 4.5) at 45 ml h^{-1} with similtaneous monitoring at 226 and 277 nm. The protein was eluted in fractions 8-12 (105-180 ml) and these were combined and analysed by HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B).

2. **Stepwise Dialysis**

The combined fractions from gel filtration were sequencially dialysed (MWCO 2000 Da, Spectropor) against nitrogen saturated solutions of 6 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5; 4 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5 and 2 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5 for 24 hours each step. As a final step the protein was dialysed against nitrogen saturated 50 mM NH_4OAc for 3 x 24 hours. The dialysed material was then lyophilised to concentrate the solution to 10-15 ml.

3. **Cation Exchange Chromatography**

Cation exchange chromatography was carried out using a CM-Sepharose (CLB beads Pharmacia) column (320 x 18 mm). Bovine ubiquitin dissolved in 50 mM NH_4OAc at pH 4.5 (15 ml) was loaded onto the CM-Sepharose column pre-equilibrated with the same buffer. A pH gradient of 50 mM NH_4OAc at pH 4.5 to 50 mM NH_4OAc at pH 5.5 was run at 30 ml h^{-1} for 19 hours with simultaneous monitoring at 226 and 277 nm. The column was then run isocratically at 50 mM NH_4OAc at pH 5.5 for 3 hours after which a salt gradient of 50 mM NH_4OAc at pH 5.5 to 0.3 M NH_4OAc at pH 5.5 was applied to elute the protein from the column. The protein was eluted in fractions 86-91 (860-920 ml, approx. 0.15 M NH_4OAc). These were combined and repeatedly lyophilised to give a white solid (78.8 mg). Analysis by HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B), amino acid analysis and IEF confirmed the material was bovine ubiquitin.

4. Anion Exchange Chromatography

Anion exchange chromatography was performed on a DEAE-Sepharose (Pharmacia) column (320 x 18 mm) equilibrated with 50 mM NH_4HCO_3 at pH 9.3. Bovine ubiquitin (78.8 mg) was dissolved in starting buffer (10 ml) and loaded onto the pre-equilibrated column. A salt gradient of 50 mM NH_4HCO_3 at pH 9.3 to 0.3 M NH_4HCO_3 at pH 9.3 was run at 30 ml h^{-1} for 21 hours with simultaneous monitoring at 226 and 277 nm. The protein was eluted in fractions 18-23 (180-240 ml, approx. 0.13 M NH_4HCO_3). These were combined and repeatedly lyophilised to give a white solid (50.6 mg). Analysis by HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B), amino acid analysis and IEF confirmed the material was homogeneous bovine ubiquitin.

Synthesis of Ubiquitin

Fmoc-Gly-OH was attached to the *p*-alkoxybenzyl alcohol resin via the amino acid chloride (0.288 mmol of Fmoc-Gly per gram of resin, 41% loading efficiency). No glycylglycine content could be detected by amino acid analysis. The synthesis of ubiquitin was carried out using the functionalised resin on a 0.25 mmol scale. The side-chains of amino acids were protected as previously indicated. Most amino acids were treble coupled, once as a symmetrical anhydride and twice as an HOBt active ester. Glycine was single coupled as a symmetrical anhydride and histidine, glutamine and asparagine were each coupled three times as their HOBt active esters. The synthesis was monitored using real time assessment of the deprotection solution in a continuous flow mode from the synthesiser to a UV detector set at 302 nm. The final Fmoc group was left on the resin bound peptide (substitution 0.032 mmol/g, 54% of theoretical maximum) which was stored in dioxan at 0-4°C until required. The resin was divided into three equal portions and each deprotected and cleaved in the same way. Note that all weights are in effect aggregate values. The peptidyl-resin was swollen in DMF then treated with 20% piperidine in DMF for 15 minutes. After washing with DMF (3 x 50 ml), DCM (3 x50 ml) and diethyl ether (3 x 50 ml) the deprotected resin was dried in a dessicator overnight. The dry resin (2.42 g) was swollen in a mixture of anisole, thioanisole and ethyl methyl sulphide (1 ml of each, 5% v:v) for 30 minutes under nitrogen before adding H_2O (1 ml, 5% v:v) and TFA (16 ml, 80% v:v). After stirring under nitrogen for the optimal time of 3.5 hours the resin was filtered off and washed with TFA (2 ml) (dry wight=623 mg). The combined washings and filtrate were then concentrated *in vacuo* and the resulting brown residue triturated with diethyl ether containing 2%

β-mercaptoethanol (50 ml) to give a white solid. This was filtered off and immediately suspended in a buffer containing 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5 (15 ml). This suspension was stirred for 16 hours to allow the crude peptide to fully dissolve.

Purification of Synthetic Ubiquitin

The synthetic material was purified in three batches in the following way.

1. **Gel Filtration Chromatography**

The solution containing the crude peptide was applied to a Sephadex G50 (fine) column (800 x 30 mm) pre-equilibrated with running buffer. The column was then eluted with 8 M urea, 10 mM DTT and 50 mM NH_4OAc (pH 4.5) at 45 ml h^{-1} with similtaneous monitoring at 226 and 277 nm. Synthetic ubiquitin was eluted in fractions 13-19 (180-285 ml). These were combined and analysed by HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B).

2. **Stepwise Dialysis**

The combined fractions from gel filtration were sequencially dialysed (MWCO 2000 Da, Spectropor) against nitrogen saturated solutions (2 l) of 6 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5; 4 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5 and 2 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5 for 24 hours each step. As a final step the protein was dialysed against nitrogen saturated 50 mM NH_4OAc for 3 x 24 hours. The dialysed material was then lyophilised to concentrate the solution to 10-15 ml. Analytical HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B) and amino acid analysis confirmed the improved purity of the synthetic material.

3. **Cation Exchange Chromatography**

The concentrated solution from dialysis was loaded onto the CM-Sepharose column (320 x 18 mm) pre-equilibrated with 50 mM NH_4OAc at pH 4.5. A pH gradient of 50 mM NH_4OAc at pH 4.5 to 50 mM NH_4OAc at pH 5.5 was run at 30 ml h^{-1} for 20 hours with simultaneous monitoring at 226 and 277 nm. The column was then run isocratically at 50 mM NH_4OAc (pH 5.5) for 3 hours after which a salt gradient of 50 mM NH_4OAc at pH 5.5 to 0.3 M NH_4OAc at pH 5.5 was applied to elute the protein from the column. The desired material was eluted in fractions 62-72 (620-730 ml, approx. 0.10 M NH_4OAc). These fractions were combined and repeatedly

lyophilised to give a white solid (125 mg). This material was analysed by HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B), amino acid analysis and IEF.

4. Anion Exchange Chromatography

The synthetic material isolated from the CM-Sepharose column was dissolved in 50 mM NH_4HCO_3 at pH 9.3 (20 ml) and applied to a DEAE-Sepharose column (320 x 18 mm) pre-equilibrated with the same buffer. A salt gradient of 50 mM NH_4HCO_3 at pH 9.3 to 0.3 M NH_4HCO_3 at pH 9.3 was run at 30 ml h^{-1} for 20 hours with simultaneous monitoring at 226 and 277 nm. The protein was eluted in two major bands labelled A and B which consisted of fractions 12-15 (120-160 ml, approx. 0.11 M NH_4HCO_3) and fractions 20-24 (200-250 ml, 0.14 M NH_4HCO_3) respectively. These were combined and repeatedly lyophilised to give 71.8 mg of A and 21.7 mg of B. Analysis by HPLC (C_8 RP300, gradient system A, R_t=16.9 min., 45% B), amino acid analysis, IEF and electrospray mass spectrometry indicated both materials were chemically identical to authentic bovine ubiquitin (see discussion section 2.4.2).

Synthesis of Ubiquitin ^{15}N-Gly ^{5,43,50,56,67}Ala ^{26,30} (Ubdes-Core)

N^{α}-9-Fluorenylmethoxycarbonyl-^{15}N-glycine
Fmoc-Gly-OH (53)

The compound was prepared by the method of Shute and Rich [215]. ^{15}N-Glycine (1 g, 13.4 mmol) was dissolved in water (10 ml) and triethylamine (4.67 ml, 33.5 mmol) added with stirring. A suspension of 9-fluorenylmethylsuccinimidyl carbonate (4.88 g, 14.5 mmol) in dioxan (10 ml) was added and mixture stirred at room temperature for 24 hours. The solution was then acidified (pH 1-2) with 2M $KHSO_4$ before being extracted with ethyl acetate (3 x 20 ml). The combined organic layers were washed with water, dried over anhydrous $MgSO_4$ and concentrated to give a yellow oil. Trituration of this oil with 40-60 petroleum gave a white solid which was recrysyallised from ethyl acetate (3.06 g, 78%); m.p. 171-173°C (lit = 173-174°C); t.l.c. R_F=0.54 (ninhydrin negative); m/z (FAB), 299 (MH^+), 257, 234, 179.

Synthesis of Ubdes-Core

Fmoc-Gly-OH was attached to the *p*-alkoxybenzyl alcohol resin via the amino acid chloride (0.291 mmol/g, 41% loading efficiency). No glycylglycine content could

be detected by amino acid analysis. The synthesis of ubiquitin was carried out using the functionalised resin Fmoc-Gly-($OCH_2C_6H_4OR$) on a 0.25 mmol scale. The side-chains of amino acids were protected as previously described for the synthesis of ubiquitin. Most amino acids were coupled once as a symmetrical anhydride followed by twice as an HOBt active ester. ^{15}N-Glycine was single coupled as an HOBt active ester and ^{14}N-glycine was single coupled as a symmetrical anhydride. Histidine, glutamine and asparagine were each coupled three times as their HOBt active esters. The synthesis was monitored using real time assessment of the deprotection solution in a continuous flow mode from the synthesiser to a UV detector set at 302 nm. The final Fmoc group was left on the resin bound peptide (substitution 0.027 mmol/g, 44% of theoretical maximum) which was stored in dioxan at 0-4^{o}C until required. The resin was divided into three equal portions and each deprotected and cleaved in the same way. Manual deprotection of the final Fmoc group was achieved by treating the resin with 20% piperidine in DMF for 15 minutes. After washing with DMF, DCM and diethyl ether the deprotected resin was dried in a dessicator overnight. The peptide was cleaved from the resin with simultaneous removal of acid-labile side-chain protection in the following way. Dry resin (2.63 g) was swollen in a mixture of anisole, thioanisole and ethyl methyl sulphide (1 ml of each, 5% v:v) for 30 minutes under nitrogen before adding H_2O (1 ml, 5% v:v) and TFA (16 ml, 80% v:v). After stirring under nitrogen for the optimum time of 4.5 hours the resin was filtered off and washed with TFA (2 ml) (dry wight=937.3 mg). The combined washings and filtrate were then concentrated *in vacuo* and the resulting brown residue triturated with diethyl ether containing 2% β-mercaptoethanol (50 ml) to give a white solid. This was filtered off and immediately suspended in a buffer containing 8 M urea, 10 mM DTT and 50 mM NH_4OAc at pH 4.5 (15 ml). This suspension was stirred for 16 hours to allow the crude peptide to fully dissolve.

Purification of Ubdes-Core

The synthetic material was purified in three batches in the following way. Note that weights given are in effect aggregate values.

1. Gel Filtration Chromatography

The solution containing the crude peptide was applied to a Sephadex G50 (fine) column (800 x 30 mm) pre-equilibrated with running buffer. The column was then eluted with 8 M urea, 10 mM DTT and 50 mM NH_4OAc (pH 4.5) at 45 ml h^{-1} with similtaneous monitoring at 226 and 277 nm. The peptide was eluted in fractions 14-

21 (195-315 ml). These were pooled and analysed by HPLC (C_8 RP300, gradient system A, R_t=15 min., 45% B) and amino acid analysis.

2. Stepwise Dialysis

The combined fractions from gel filtration were sequencially dialysed (MWCO 2000 Da, Spectropor) against nitrogen saturated solutions (2 l) of 6 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5; 4 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5 and 2 M urea, 10 mM DTT, 50 mM NH_4OAc at pH 4.5 for 24 hours each step. As a final step the protein was dialysed against nitrogen saturated 50 mM NH_4OAc for 3 x 24 hours before being lyophilised to concentrate the solution to 10-15 ml ready for cation exchange chromatography.

3. Cation Exchange Chromatography

The concentrated solution from dialysis was applied to a CM-Sepharose column (320 x 18 mm) pre-equilibrated with 50 mM NH_4OAc at pH 4.5. A pH gradient of 50 mM NH_4OAc at pH 4.5 to 50 mM NH_4OAc at pH 5.5 was run at 30 ml h^{-1} for 20 hours with simultaneous monitoring at 226 and 277 nm. The column was then run isocratically at 50 mM NH_4OAc at pH 5.5 for 5 hours after which a salt gradient of 50 mM NH_4OAc at pH 5.5 to 0.3 M NH_4OAc at pH 5.5 was applied to elute the protein from the column. The desired material was eluted in fractions 100-110 (1000-1110 ml, approx. 0.14 M NH_4OAc). These fractions were combined and repeatedly lyophilised to give a white solid (155 mg) which was analysed by HPLC (C_8 RP300, gradient system A, R_t=15 min., 40% B), amino acid analysis and IEF.

4. Anion Exchange Chromatography

The synthetic material isolated from the CM-Sepharose column was dissolved in 50 mM NH_4HCO_3 at pH 9.3 (15 ml) and loaded onto a DEAE-Sepharose column (320 x 18 mm) pre-equilibrated with the same buffer. The column was then run isocratically at 50 mM NH_4HCO_3 at pH 9.3 for 3 hours at 30 ml h^{-1} with simultaneous monitoring at 226 and 277 nm afterwhich a salt gradient of 50 mM NH_4HCO_3 at pH 9.3 to 0.3 M NH_4HCO_3 at pH 9.3 was applied to elute the peptide. The desired material was eluted in fractions 33-38 (330-390 ml, 0.17 M NH_4HCO_3) These were combined and repeatedly lyophilised to give a white solid (92 mg). which was analysed by HPLC (C_8 RP300, gradient system A, R_t=15 min., 40% B), amino acid analysis and IEF.

5. Semi-Preparative HPLC

The synthetic material was further purified by semi-preparative HPLC on an Aquapore C_8 reverse-phase column (100 x 10 mm, 20 μm, gradient system A) using a flow rate of 5 ml/min. Removal of the solvents by lyophilisation gave the *title* compound (81 mg, 3.9%) as judged by HPLC (C_8 RP300, gradient system A, R_t=15 min., 40% B), amino acid analysis, IEF and electrospray mass spectrometry (see section 2.5.2).

Synthesis of Human 52 Residue Carboxy-Extension Protein (CEP52)

Fmoc-Lys(Boc)-OH was attached to the *p*-alkoxybenzyl alcohol resin via the symmetrical anhydride (0.104 mmol/g, 14% loading efficiency). The synthesis was carried out on the functionalised resin Fmoc-Lys(Boc)-($OCH_2C_6H_4OR$) on a 0.10 mmol scale. The side-chains of amino acids were protected as previously described. All five cysteines were protected with the S-Bu^t group. Most amino acids were coupled once as a symmetrical anhydride followed by twice as an HOBt active ester. Exceptions to this were glycine which was single coupled as a symmetrical anhydride and histidine, glutamine and asparagine which were each coupled three times as their HOBt active esters. All Fmoc deprotections were monitored in a continuous flow mode from the synthesiser to a UV detector set at 302 nm. The final Fmoc group was left on the resin-bound peptide, two thirds of which were stored for later use. Manual deprotection of the final Fmoc group was accomplished by treating the resin with 20% piperidine in DMF for 20 minutes. After washing well with DMF, DCM and diethyl ether the deprotected resin was dried in a dessicator overnight. The peptide was cleaved from the resin with simultaneous removal of side-chain protection (with the exception of the cysteine protection) in the following way. Dry resin (634.5 mg) was swollen in a mixture of anisole, phenol, ethanedithiol and ethyl methyl sulphide (1 ml of each, 5% v:v) for 30 minutes under nitrogen before adding H_2O (1 ml, 5% v:v) and TFA (15 ml, 75% v:v). After stirring under nitrogen for the optimum time of 6 hours the resin was filtered off and washed with TFA (2 ml) (dry weight=335.8 mg). The combined washings and filtrate were then concentrated *in vacuo* and the crude peptide (299 mg) was precipitated with diethyl ether. This was filtered off, dissolved in 30% aqueous acetic acid (20 ml) and applied to a Sephadex G50 column (90 x 2.5 cm) pre-equilibrated with the same buffer. The column was eluted with 30% acetic acid at 30 ml h^{-1} with monitoring at 226 and 277 nm. The peptide was eluted in fractions 7-11 (70-120 ml). These fractions were combined and sequencially

dialysed (MWCO 2000 Da) against solutions (2 l) of 30% AcOH; 20% AcOH; 10% AcOH and H_2O for 24 hours each step. Following lyophilisation the cysteine t-butylthio protection was removed by Bu_3P treatment. The cysteine-protected peptide was dissolved in TFE (5 ml), H_2O (0.25 ml) and Bu_3P (0.3 ml) was added. The reaction was left stirring under nitrogen for 3 hours after which the solvent was removed *in vacuo* and the peptide filtered off and washed with diethyl ether. The fully deprotected peptide was immediately dissolved in a buffer containing 10 mM DTT and 50 mM NH_4OAc at pH 6.0 (15 ml) and subjected to a second round of gel filtration on a Sephadex G50 column (90 x 2.5 cm) pre-equilibrated with the same buffer. Elution of the column with 10 mM DTT, 50 mM NH_4OAc (pH 6.0) at 30 ml h^{-1} afforded the desired material in fractions 7-13 (70-140 ml). These fractions were combined and concentrated (20-30 ml) by lyophilisation.

The peptide was purified by semi-preparative HPLC on an Aquapore C_8 reverse-phase column (100 x 10 mm, 20 μm, gradient system B) using a flow rate of 5 ml/min. Removal of the solvents by lyophilisation gave the *title* compound (21 mg, 3.4%) as judged by HPLC (C_8 RP300, gradient system A, R_t=14.5 min., 38% B), amino acid analysis and electrospray mass spectrometry (see section 2.6.3).

Synthesis of Human Ubiquitin (^{15}N-Gly76)- 52 Residue Carboxy-Extention Protein (UbCEP52)

The synthesis of UbCEP52 was carried out using the remaining two thirds of the Fmoc protected resin-bound CEP52. In effect the construction of UbCEP52 was a continuation of the initial CEP52 synthesis. The side-chains of amino acids were protected as previously described for ubiquitin. Most amino acids were coupled once as a symmetrical anhydride followed by twice as an HOBt active ester. ^{15}N-glycine-76 was single coupled as an HOBt active ester, all other glycines were were single coupled as symmetrical anhydrides. Histidine, glutamine and asparagine were each coupled three times as their HOBt active esters. All Fmoc deprotections were monitored in a continuous flow mode from the synthesiser to a UV detector set at 302 nm. The final Fmoc group was left on the resin-bound peptide (substitution 0.0135 mmol/g, 73% of theoretical maximum). Manual deprotection of the final Fmoc group was accomplished by treating the resin with 20% piperidine in DMF for 20 minutes. After washing well with DMF, DCM and diethyl ether the deprotected resin was dried in a dessicator overnight. The peptide was cleaved from the resin with simultaneous removal of side-chain protection

(with the exception of the cysteine protection) in the following way. Dry deprotected resin (768.1 mg) was swollen in a mixture of anisole, phenol, ethanedithiol and ethyl methyl sulphide (1 ml of each, 5% v:v) for 30 minutes under nitrogen before adding H_2O (1 ml, 5% v:v) and TFA (15 ml, 75% v:v). After stirring under nitrogen for the optimum time of 6 hours the resin was filtered off and washed with TFA (2 ml) (dry weight=205.5 mg). The combined washings and filtrate were then concentrated *in vacuo* and the crude peptide (562.6 mg) was precipitated with diethyl ether.The synthetic protein was then suspended in a buffer containing 8 M urea and 50 mM NH_4OAc at pH 6.0 (20 ml) and stirred overnight to allow all the material to dissolve. The crude protein solution was then applied to a Sephadex G75 (superfine) column (90 x 2.5 cm) pre-equilibrated with running buffer and eluted with 8 M urea, 50 mM NH_4OAc (pH 6.0) at 30 ml h^{-1} with monitoring at 226 and 277 nm. The peptide was eluted in fractions 32-53 (320-540 ml). These fractions were combined and sequencially dialysed (MWCO 10000 Da) against solutions (2 l) of 6 M urea, 50 mM NH_4OAc (pH 6.0); 4 M urea, 50 mM NH_4OAc (pH 6.0) and 2 M urea, 50 mM NH_4OAc (pH 6.0) for 24 hours each step. As a final step the protein was dialysed against nitrogen saturated H_2O for 3 x 24 hours before being lyophilised. The cysteine t-butylthio protection was removed by Bu_3P treatment. The cysteine-protected peptide was dissolved in TFE (10 ml), H_2O (0.50 ml) and Bu_3P (0.50 ml) was added. The reaction was left stirring under nitrogen for 3 hours afterwhich the solvent was removed *in vacuo* and the peptide filtered off and washed with diethyl ether. The fully deprotected peptide was immediately dissolved in a buffer containing 10 mM DTT, 50 mM NH_4OAc at pH 6.0 (20 ml) and subjected to a second round of gel filtration on a Sephadex G50 column (90 x 2.5 cm) pre-equilibrated with running buffer. Elution of the column with 10 mM DTT, 50 mM NH_4OAc (pH 6.0) at 30 ml h^{-1} afforded the desired material in fractions 5-12 (50-130 ml). These fractions were combined and repeatedly lyophilised.

Initial small scale attempts at cation exchange were as follows.

1. CM-Sepharose Cation Exchange Chromatography

The fully deprotected synthetic material (5 mg) was dissolved in a buffer containing 10 mM DTT and 50 mM NH_4HCO_3 at pH 9.5 (2 ml) and applied to CM-Sepharose column (150 x 13 mm) previously equilibrated with the same buffer. A salt gradient of 10 mM DTT, 50 mM NH_4HCO_3 at pH 9.5 to 10 mM DTT, 0.5 M NH_4HCO_3 at pH 9.5 was run at 15 ml h^{-1} for 25 hours with

monitoring at 226 and 277 nm. This was followed by a second salt gradient of 0.5 M NH_4HCO_3 (pH 9.5), 10 mM DDT to 1 M NH_4HCO_3 (pH 9.5), 10 mM DTT. No protein was observed to elute from the column during either of these gradients.

2. **Analytical Cation Exchange HPLC**

Cation exchange HPLC was performed on an Aquapore CX-300 column (100 x 4.6 mm, 7 μm) employing a flow rate of 1 ml/min and monitoring at 280 nm. The protein dissolved in starting buffer was applied (30 μl of a 1 mg/ml solution) to the column and a linear gradient of 0-100% B over 30 minutes carried out using each of the systems outlined below.

(i) A = 50 mM NH_4OAc at pH 4.0; B = 1 M NaCl, 50 mM NH_4OAc at pH 4.0.

Synthetic material - did not elute form column.
Bovine ubiquitin - R_t = 9 min., 30% B.

(ii) A = 50 mM MOPS at pH 6.0; B = 1 M NaCl, 50 mM MOPS at pH 6.0.

Synthetic material - did not elute from column.
Bovine ubiquitin - R_t = 4.5 min., 15% B.

(iii) A = 50 mM HEPES at pH 8.0; B = 1 M NaCl, 50 mM Hepes at pH 8.0.

Synthetic material - did not elute from column.
Bovine ubiquitin - R_t = 0 min., 0% B.

(iv) A = 50 mM Tricine at pH 9.5; B = 1 M NaCl, 50 mM Tricine at pH 9.5.

Synthetic material - did not elute from column.

(v) A = 50 mM Bis-tris propane at pH 11.0; B = 1 M NaCl, 50 mM Bis-tris propane at pH11.0.

Synthetic material - partially soluble at this pH, did elute as a broad band
R_t = 24 min., 80% B.

3. **Heparin-Agarose Weak Cation Exchange Chromatography**

The synthetic protein (4 mg) was dissolved in a buffer containing 50 mM Tris.HCl at pH 8.0, 50 mM $(NH_4)_2SO_4$, 1 mM DTT and 1 mM $ZnCl_2$ (5 ml) and applied to a heparin-agarose column (80 x 13 mm) pre-equilibrated with the same buffer. A salt gradient of 50 mM Tris.HCl at pH 8.0, 50 mM $(NH_4)_2SO_4$, 1 mM DTT and 1 mM

$ZnCl_2$ to 50 mM Tris.HCl at pH 8.0, 600 mM $(NH_4)_2SO_4$, 1 mM DTT and 1 mM $ZnCl_2$ was run at 15 ml h^{-1} for 22 hours. This was followed by a second salt gradient of 50 mM Tris.HCl at pH 8.0, 600 mM $(NH_4)_2SO_4$, 1 mM DTT and 1 mM $ZnCl_2$ to 2 M NaCl, 50 mM Tris.HCl at pH 8.0, 600 mM $(NH_4)_2SO_4$, 1 mM DTT and 1 mM $ZnCl_2$. HPLC analysis of the resulting fractions indicated no protein had eluted from the column during either gradient.

Cation exchange chromatography was successfully carried out using the following system.

4. CM-Sepharose Cation Exchange Chromatography

The synthetic material (56 mg) was dissolved in a buffer containing 50 mM NH_4Cl, 1 mM DTT, 0.1 mM $ZnCl_2$ and 50 mM formic acid at pH 4.0 and applied to a CM-Sepharose column (320 x 18 mm) previously equilibrated with the same buffer. A salt gradient of 50 mM NH_4Cl, 1 mM DTT, 0.1 mM $ZnCl_2$, 50 mM formic acid at pH 4.0 to 1.5 M NH_4Cl, 1 mM DTT, 0.1 mM $ZnCl_2$, 50 mM formic acid at pH 4.0 was run at 30 ml h^{-1} for 30 hours with simultaneous monitoring at 226 and 277 nm. The desired material was eluted in fractions 43-47 (430-480 ml, approx. 1.1 M NH_4Cl). These fractions were combined and dialysed (MWCO 2000 Da) against water for 24 hours after which lyophilisation afforded a white solid (16 mg, 1.1%). Analysis by HPLC (C_8 RP300, gradient system A, R_t=15.7 min., 42% B), amino acid analysis, SDS-PAGE, sequencing and laser desorption mass spectrometry indicated this was the required material (see section 2.6.5).

Isoelectric Focussing Electrophoresis and Western Blotting

All samples were dissolved in 50 mM Tris., pH 8.0 (1 mg/ml) of which 20 µl aliquots were loaded onto a pre-focussed broad range pH 3.5-9.5 PAG plate gel (Pharmacia). Electrofocussing was then carried out at 18 mA, 5 W for 1000 Vh after which the gel was either Coomassie stained or prepared for electroblotting by placing in a transfer buffer containing 250 mM Tris., pH 8.3, 1.5 M glycine and 20% v:v MeOH. After removal of the plastic backing from the gel the protein was electroeluted onto nitrocellulose for 2 h at 0.6 to 1.2 mA (LKB 2005 Transphor system).

Following electroblotting the remaining sites on the nitrocellulose were blocked overnight in 10 mM Tris. pH 8.0, 400 mM NaCl (TBS) containing 20% by weight

dried milk. After washing with TBS (3 x 5 min) the nitrocellulose was placed in TBS containing 5% w:v dried milk (30 ml) with rabbit anti-ubiquitin IgG polyclonal antibodies (5 µl) and shaken overnight. The nitrocellulose was then washed with TBS (3 x 5 min) before being treated with TBS (30 ml) containing anti-rabbit IgG horse-raddish peroxidase conjugated secondary antibodies (10 µl, Sigma) and shaken for 2 hours. After washing well with TBS (3 x 10 min) the secondary antibody was visualized by staining with a mixture of dianisidine (0.5 ml, 5 mg/ml), imidazole (1 mg, 0.1 M pH 7.4), 30% H_2O_2 and H_2O (8.4 ml).

Total Enzymic Digestion

Protein (1 mg) was dissolved in performic acid (100 µl) at 0°C for 1 hour after which the solution was lyophilised and the resulting residue redissolved in 8 M urea, 5% formic acid (100 µl) and left for 2 h. This solution was dialysed (MWCO 500 Da) against 5% formic acid over 24 hours before being lyophilised. The resulting solid was taken up in 5% formic acid (100 µl) and digested with pepsin (1% by wt, 25 µl) for 24 hours after which a fresh aliquot of pepsin was added and the digestion allowed to proceed for a further 24 h. The solution was lyophilised and dissolved in 0.2 M NH_4HCO_3 at pH 7.75 (0.25 ml) with leucine aminopeptidase microsomal (2 U/ml, 15 µl), prolidase (200 U/ml, 28.1 µl) and 25 mM $MnCl_2$ (50 µl). The digestion was allowed to proceed for 48 h after which the solution was repeatedly lyophilised and the resulting residue was dissolved in pH 2.2 citrate buffer (2 ml) for amino acid analysis.

Tryptic Mapping

Tryptic digestion was carried out according to the method of Cox *et al* [196]. Protein (5 mg) was dissolved in 1 ml of a buffer containing 50 mM Tris. (pH 8.2), 6.5 M urea and 10 mM $CaCl_2$. Trypsin (5% w:w with the protein) was added and the digestion allowed to proceed for 4 hours at 37°C before a further aliquot of trypsin (5% w:w) was added. After a total of 8 h digestion the protein solution was acidified to pH 2 with 6 M HCl and stored at -20°C until required.

Peptide mapping was carried out with reverse-phase HPLC using a C_{18} column (Vydac, 5 µm, 250 x 4.1 mm) and the following solvent system.

A: 0.1% TFA, 0.1% TEA in H_2O

B: 0.1% TFA, 0.1% TEA in 40% aqueous acetonitrile

The digest solution was injected onto the column in 20 µl aliquots and a linear gradient employed with a flow rate of 1 ml/min. The gradient began at 10% B for 2 min, rose to 70% B at 50 min, and then to 100% B at 52.5 min. It was then left to run isocratically at 100% B for 5 min and then re-equilibrated to 10% B. UV monitoring was carried out at 205 nm (1.5 Aufs) and the peptide fragments were isolated in 1 ml fractions. These were lyophilised, dissolved in citrate buffer at pH 2.2 and subjected to amino acid analysis for identification.

Results obtained for Synthetic Ubiquitin

Peptide	R_t (min)	Amino Acid Analysis
T0 (28-29)	3.2	Ala_1 1.0 :Lys_1 1.1
T1 (30-33)	4.4	Asx_1 1.0 :Glx_1 1.1 :Ile_1 0.9 :Lys_1 1.0
T2 (73-76)	5.4	Gly_2 1.8 :Leu_1 1.12 :Arg_1 0.9
T3 (7-11)	8.8	Thr_2 1.9 :Gly_1 1.0 :Leu_1 1.0 :Lys_1 1.0
T4 (49-54)	11.0	Asx_1 1.0 :Glx_2 2.2 :Gly_1 1.0 :Leu_1 0.9 :Arg_1 0.9
T5 (34-42)	16.6	Asx_1 1.2 :Glx_3 2.7 :Gly_1 1.0 :Pro_2 2.3 :Ile_1 0.9 :Arg_1 0.9
T6 (55-63)	28.0	Asx_2 2.1 :Thr_1 1.0 :Ser_1 0.9 :Glx_1 1.4 :Ile_1 1.0 :Leu_1 1.0 :Tyr_1 0.9 :Lys_1 1.0
T7 (43-48)	34.0	Gly_1 1.2 :Ala_1 1.0 :Ile_1 0.9 :Leu_1 1.0 :Phe_1 0.9 :Lys_1 1.0
T8 (1-6)	37.6	Glx_1 1.3 :Val_1 1.0 :Met_1 0.7 :Ile_1 0.9 :Phe_1 0.9 :Lys_1 0.9
T9 (12-27)	42.4	Asx_2 2.2 :Thr_3 2.6 :Ser_1 1.0 :Glx_3 3.6 :Pro_1 1.3 :Val_2 2.0 :Ile_2 1.9 :Leu_1 1.1 :Lys_1 1.0
T10 (64-72)	43.6	Thr_1 1.0 :Ser_1 1.0 :Glx_1 1.6 :Val_1 1.2 :Leu_3 2.8 :His_1 1.0 :Arg_1 1.0

Figures given in parenthesis are residue numbers

Results for Ubdes-Core

Peptide	R_t (min)	Amino Acid Analysis
T0 (28-29)	4.2	Ala_1 1.1 :Lys_1 1.1
T1 (49-54)	4.2	Asx_1 1.1 :Glx_2 2.2 :Gly_2 2.8 :Arg_1 0.9
T2 (7-11)	9.3	Thr_2 1.7 :Gly_1 1.3 :Leu_1 0.9 :Lys_1 1.0
T3 (30-42)	16.5	Asx_2 2.6 :Glx_4 4.2 :Gly_1 1.7 :Pro_2 1.8 Ala_1 1.0 :Ile_1 1.0 : Lys_1 1.0 :Arg_1 0.9
T4 (55-63)	19.0	Asx_2 2.2 :Thr_1 0.9 :Ser_1 0.9 :Glx_1 1.6 Gly_1 1.1 :Ile_1 0.8 :Tyr_1 0.5 :Lys_1 1.0
T5 (43-48)	27.3	Gly_2 2.0 :Ala_1 1.0 :Ile_1 0.9 :Phe_1 1.0 :Lys_1 1.1
T6 (64-72)	28.8	Thr_1 0.9 :Ser_1 0.9 :Glx_1 1.2 Gly_1 1.3 :Val_1 0.8 :Leu_2 2.3 :His_1 0.9 :Arg_1 1.0
T7 (1-6)	33.0	Glx_1 1.5 :Gly_1 1.3 :Met_1 - :Ile_1 1.0 :Phe_1 0.9 :Lys_1 1.1
T8 (12-27)	38.0	Asx_2 2.4 :Thr_3 3.1 :Ser_1 1.0 :Glx_3 3.0 :Pro_1 1.3 :Ala_1 1.0 : Val_1 0.9:Ile_2 1.7 :Leu_1 1.1 :Lys_1 1.2

Figures given in parenthesis are residue numbers

Crystallization of Synthetic Ubiquitin

Crystals of purified synthetic ubiquitin were reproducibly obtained from a vapour diffusion (hanging drop) experiment in which the synthetic solution was seeded with a natural ubiquitin crystal. Several large natural ubiquitin crystals were supplied by Dr W.J. Cook (University of Alabama Medical School). One such crystal was carefully added to a single 10 µl drop containing 5 µl of concentrated synthetic ubiquitin solution (20 mg/ml in H_2O) and 5 µl from the well solution which contained 30% (w:v) polyethylene glycol 4000 (PEG4000) in 50 mM cacodylate-HCl (pH 5.6). Small single crystals formed independently of the seed crystal in the original drop and were transferred into fresh drops of the same composition for further growing. These crystals were then removed from the second drop and put into fresh ones containing PEG4000 at a concentration of 25% (w:v). This PEG concentration was gradually increased to 38% in order to stabilize the crystals.

Crystallization of Ubdes-Core

Attempts were made at the crystallization of purified Ubdes-Core without initial seeding using the vapour diffusion (hanging drop) technique. In each case the 10 µl hanging drop contained 5 µl of concentrated Ubdes-Core solution (10 mg/ml in H_2O) and 5 µl from the well solution. The following 18 well solutions were screened:

50 mM cacodylate-HCl (pH 5.0) containing 36, 38 and 40% (w:v) PEG4000
50 mM cacodylate-HCl (pH 5.2) containing 36, 38 and 40% (w:v) PEG4000
50 mM cacodylate-HCl (pH 5.4) containing 36, 38 and 40% (w:v) PEG4000
50 mM cacodylate-HCl (pH 5.6) containing 36, 38 and 40% (w:v) PEG4000
50 mM cacodylate-HCl (pH 5.8) containing 36, 38 and 40% (w:v) PEG4000
50 mM cacodylate-HCl (pH 6.0) containing 36, 38 and 40% (w:v) PEG4000

No crystals of Ubdes-Core were grown under any of the above conditions.

Fast Protein Liquid Chromatography (FPLC) Studies on the Partial Unfolding of Bovine and Synthetic Ubiquitin

Protein was dissolved in 20 mM piperazine at pH 9.6 (1 mg/ml, 0.116 mM) and incubated at 25°C. Aliquots (100 µl) of this solution were periodically analysed by FPLC using a Mono Q anion exchange column (Pharmacia, HR 5/5, 10 µm) and the following solvent system:

A = 20 mM piperazine at pH 9.6.
B = 20 mM piperazine at pH 9.6, 1 M NaCl.

The protein was eluted from the column using a linear salt gradient with a flow rate of 0.5 ml/min. The gradient began at 0% B for 6 min, rose to 40% B at 70 min, and then to 100 % B at 76 min before being re-equilibrated at 0% B. UV monitoring was carried out at 280 nm (Aufs=0.05). R_t of native ubiquitin = 21 min., 12% B and R_t of denatured ubiquitin = 22 min., 13.5% B.

Spectrophotometric Titration of the Phenolic Group in Tyrosine

Measurements of pH were made on a Corning Ion Analyzer 150 calibrated with standard buffer solutions (Whatman) at pH 7.0 and 10.0. All measurements were

accurate to ± 0.01 of a pH unit. UV absorption measurements were carried out using 3 ml quartz cuvettes of light path 10 mm. The temperature of the cell was maintained at 25 or 35°C by circulating water from a thermostatic bath at the appropriate temperature. All $\varepsilon_{295\ nm}$ values are accurate to within ± 75 $dm^3\ mol^{-1}\ cm^{-1}$.

Protein was dissolved in 0.16 M KCl solution (0.5 mg/ml) and the pH altered to around 7.0 by the careful addition of the appropriate amount of 0.1 M HCl or 0.5 M KOH. The UV absorbance of this solution was then measured in the range 400-229 nm (Aufs=0.5). The pH of the protein solution was then slowly increased in increments of around 0.5 by addition of 0.5 M KOH and at each stage a UV spectrum recorded. The temperature of the protein solution was maintained at either 25 or 35°C between successive UV measurements. All absorbance measurements were corrected for the gradual dilution of the protein solution caused by addition of the base. The data are presented as plots of $\varepsilon_{295\ nm}$ *vs.* pH.

Circular Dichroism Spectroscopy

Spectra were recorded in the region 240-190 nm at 25°C. Samples were dissolved in 50 mM boric acid (pH 6.0 or 7.0 as indicated in the text) containing the additives referred to in the text. Protein concentrations were typically 50 μM.

Fluorescence Quenching Studies

Solutions of protein (50 μM) in 50 mM NH_4OAc (pH 6.0), 0.16 M KCl were incubated for 45 minutes (25°C) with 0-1 M KI. $Na_2S_2O_3$ was added to a concentration of 0.1 mM to prevent formation of I^{3-}. Fluorescence emission was monitored at 25°C at 303 nm with excitation at 280 nm (Em./Ex. slit width = 5 nm) and expressed as F^o/F where F^o is fluorescence of the protein in the absence of quencher. The data are presented as direct Stern-Volmer plots of F^o/F *vs.* quencher concentration.

Chemical Denaturation Studies

All experiments used highly purified enzyme grade guanidinium hydrochloride (GdmCl) purchased from Life Technologies Inc. The buffer used in all denaturation experiments was 4-morpholinepropanesulphonic acid (MOPS) from Sigma. The MOPS sock solution was made to 50 mM at pH 7.0 using Milli-Q HPLC grade water. The guanidinium hydrochloride stock solution was prepared gravimetrically

in volumetric flasks with MOPS buffer and stored at room temperature. The molarity of the GdmCl stock solution was determined using refractive index measurements and the following equation [228]:

$$\text{Molarity} = \{57.147 \times (\Delta N)\} + \{38.68 \times (\Delta N)^2\} - \{91.60 \times (\Delta N)^3\}$$

where ΔN is the difference in refractive index between the denaturant solution and the buffer. All protein stock solutions (100 μM) were prepared with MOPS stock buffer. For each data point in the chemical denaturation experiments, 100 μl of protein stock solution was added to the appropriate amount of GdmCl stock solution and the final volume made to 1 ml with MOPS stock solution. All volumes were measured using Gilson micropipettes (accuracy ± 0.8%). The final concentration of protein solution was 10 μM. All protein/denaturant solutions were preequilibrated at 25°C for 1-2 hours before any fluorescence measurements were made. Fluorescence emission was monitored at 303 nm with excitation at 280 nm (Em./Ex. slit width = 5 nm). Thermostated cuvette holders at 25°C were used to maintain the temperature during all spectroscopic measurements. All fluorescence measurements were reproducible to within ± 1 of a fluorescence unit. The data are presented as unfolding curves of fluorescence *vs.* denaturant concentration. All unfolding curves were found to be reversible by diluting samples of unfolded protein to low GdmCl concentration and measuring the recovery of fluorescence.

Ellmans Assay for Free Thiol Groups

CEP52 or UbCEP52 were reduced by treatment with DTT. The peptide was dissolved in a buffer containing 50 mM Tris., pH 7.2, 10 mM DTT, 10 mM EDTA and sonicated overnight at 40°C. The reduced peptide was purified by semi-preparative HPLC (C_8 RP300, 100 x 10 mm, 20 μm, gradient system A), lyophilised and stored at -20°C.

The number of free thiols was determined by the reaction with 5,5'-dithiobis(*o*-nitrobenzoic acid) (DTNB) [222]. The peptide (0.5-2.0 mg) was dissolved in 2.5 ml denaturant buffer containing 0.1 M Tris. at pH 8.0, 8 M urea and 0.01 M EDTA. A solution containing 10 mM DTNB in 50 mM sodium phosphate at pH 7.0 (100 μl) was then added and the absorbance of the solution measured at 412 nm.

CEP52

Immediately after reduction: requires 9.70×10^{-6} mol dm^{-3}
found 9.17×10^{-6} mol dm^{-3}

After 1 hour in water: requires 2.20×10^{-4} mol dm^{-3}
found 1.03×10^{-4} mol dm^{-3}

After 24 hours in water: requires 1.16×10^{-4} mol dm^{-3}
found 2.66×10^{-5} mol dm^{-3}

UbCEP52

requires 7.05×10^{-5} mol dm^{-3}
found 1.89×10^{-6} mol dm^{-3}

Cobalt (II) Binding Studies

CEP52 or UbCEP52 were reduced by treatment with DTT. The peptide was dissolved in a buffer containing 50 mM Tris., pH 7.2, 10 mM DTT, 10 mM EDTA and sonicated overnight at 40°C. The reduced peptide was purified by semi-preparative HPLC (C_8 RP300, 100 x 10 mm, 20 μm, gradient system A), lyophilised and stored at -20°C.

Lyophilised peptide was dissolved in a buffer containing 50 mM Tris. (pH 7.2) and 100 mM NaCl. The peptide concentration was 50 μM in the case of CEP52 and 25μM for UbCEP52. The UV spectrum of the peptide solution in the region 700-250 nm (Aufs=0.5) was then recorded. $CoCl_2$ solution (10 μl, 1 equiv) was then added and the UV spectrum of the peptide solution recorded. This was repeated for 5 equivalents (50 μl) of $CoCl_2$ solution. In each case the spectrum of free $CoCl_2$ was subtracted from the spectrum of the peptide solution.

Appendix 1 - ^{1}H NMR Assignments of Synthetic Ubiquitin

^{1}H NMR assignments of synthetic mammalian ubiquitin. 20 mg ml^{-1} protein in 90% H_2O: 10% D_2O at pH 4.8. Spectrum acquired at 50°C on the 600 MHz spectrometer. δ (OH, H_2O) ≡ 4.8 ppm.

	N^{α}	α	β	β'	γ	γ'	δ	δ'	ε	N^{ω}
M1		4.26	2.30	N	N	N				
Q2	8.99	5.43	2.17	1.88	2.42	2.42				
I3	8.54	4.36	N		0.82	N	N	N		
F4	8.77	5.80	3.24	3.06	7.08		7.27		7.42	
V5	9.46	5.03	2.10		0.92	0.88				
K6	9.17	5.45	1.91	1.88	1.60	1.48	1.77	1.63	3.09	
T7	8.89	5.15	4.97		1.37					
L8	9.12	4.52	1.99	N	2.11		1.26	1.19		
T9	7.79	4.63	4.79		1.48					
G10	8.00	4.55	3.80							
K11	7.44	4.56	1.98	1.98	1.60	1.44	1.90	1.85	3.12	
T12	8.70	5.28	4.14		1.27					
I13	9.71	4.70	2.07		1.31	N	0.92	1.06		
T14	8.78	5.15	4.23		1.32					
L15	8.90	4.94	1.54	1.44	1.62		0.94	0.89		
E16	8.25	5.08	N	N	N	N				
V17	9.08	4.88	2.54		0.93	0.66				
E18	8.86	5.26	1.97	2.38	2.61	2.50				
P19		4.34	2.61	2.22	2.41	2.25	4.21	4.00		
S20	7.23	4.57	4.35	3.98						
D21	8.21	4.89	3.15	2.72						
T22	8.02	5.12	4.98		1.46					
I23	8.67	3.83	2.64		0.98	N	N	N		
E24	N	4.10	2.27	2.23	2.60	2.54				
N25	8.09	4.74	3.38	3.07						N
V26	8.27	3.60	2.54		1.18	0.89				
K27	N	N	N	N	N	N	N	N	N	
A28	8.18	4.36	1.83							
K29	8.06	4.39	2.34	2.32	1.78	1.68	2.12	1.99	N	
I30	8.44	3.70	2.54		0.87	0.87	1.07	N		
Q31	8.73	4.02	2.68	2.16	2.47	2.47				N
D32	8.14	4.55	3.04	2.94						
K33	7.65	4.52	2.21	2.21	1.91	1.80	2.04	2.04	N	
E34	8.89	N	2.44	1.88	2.35	2.26				
G35	8.65	4.33	4.13							
I36	6.37	4.63	1.60		1.28	1.13	0.98	1.65		
P37		4.82	2.62	2.20	2.27	2.27	4.37	3.77		
P38		4.32	2.42	2.27	2.42	1.86	N	3.97		
D39	8.68	4.63	2.96	2.88						
Q40	7.98	4.67	2.02	2.02	2.60	2.60				N
Q41	7.63	4.42	2.13	2.10	2.72	2.72				N
R42	8.67	4.68	1.88	1.83	1.70	1.61	3.32	3.26		7.33
L43	8.95	5.55	1.77	1.67	1.37		0.99	0.95		
I44	9.32	5.12	1.95		N	0.88	0.88	N		

F45	9.04	5.35	3.19	3.01	7.67		7.55		7.73	
A46	8.94	3.89	1.10							
G47	8.14	4.31	3.64							
K48	8.19	N	2.10	2.10	1.72	1.72	2.06	2.06	3.38	
Q49	8.73	N	2.44	2.15	2.19	2.19				N
L50	8.71	4.27	N	1.22	1.65		0.71	0.06		
E51	8.47	4.68	2.42	2.15	2.62	2.52				
D52	8.25	4.57	2.84	2.72						
G53	9.50	4.39	4.54							
R54	7.63	4.94	2.43	2.43	2.29	1.86	3.35	3.28		7.27
T55	9.02	5.42	N		1.34		N	N		
L56	8.36	4.26	2.29	1.45	N		N	N		
S57	8.61	4.45	4.05	3.96						
D58	8.10	4.49	3.17	2.47						
Y59	7.46	4.87	3.65	2.73	7.32		7.03			
N60	8.34	4.55	3.50	2.99						N
I61	7.39	3.59	1.59		0.68	0.58	0.68	0.04		
Q62	7.72	4.68	2.45	2.09	2.57	2.52				N
K63	8.57	4.17	2.23	2.10	1.68	1.68	1.92	1.92	3.26	
E64	9.43	3.52	2.59	N	2.42	N				
S65	7.84	4.86	4.10	3.85						
T66	8.79	5.48	4.25		1.16					
L67	9.57	5.27	1.83	1.82	1.96		0.87	0.87		
H68	9.39	5.35	3.06	3.24			7.01		8.02	
L69	8.49	5.36	1.83	1.28	1.54		1.05	0.94		
V70	9.31	4.56	2.24		1.04	1.13				
L71	8.13	5.24	1.84	1.73	1.88		1.16	1.05		
R72	8.74	4.47	1.75	1.96	1.70	1.70	3.34	3.34		7.27
L73	8.38	4.60	1.85	1.78	1.65		1.12	1.08		
R74	8.49	4.50	2.07	1.98	1.86	1.82	3.42	3.38		7.41
G75	8.52	4.17	4.17							
G76	8.02	4.00	3.93							

N - not observed

Appendix 2 - X-Ray Crystallography

Data Collection

A data set to a maximum resolution of 1.6 A^{o} was collected from one crystal using a Siemens-Nicholet-Xentronic area detector on a Rigaku rotating anode generator RU200 operating at 40 kV, 100 mA and fitted with a graphite monochromator to select Cu-Kα radiation. There were 14404 observations representing 6322 unique reflections to 1.6 A^{o}. The data were reduced using the Xengen v.1.3 program to give an unweighted absolute value for the R-merge on intensity of 5.3%. Because of the poor quality of the weak reflections at high resolution, the data were restricted to 1.8 A^{o} resolution thus providing 5058 out of a possible 6050 reflections which represents 85% of the expected diffraction pattern. After a 1σ(I) cutoff was applied, 4349 reflections in the range between 6.0 and 1.8 A^{o} were left for model refinement. In comparison, the final refinement of the original model of the natural ubiquitin was based upon 5554 reflections in the same range with a similar σ(I) cutoff, indicating that the weak reflections in that study were measured more accurately. In order to minimize the possibility that significant differences in structure were the result of different ambient conditions between Edinburgh and Alabama, data was collected from a crystal of the natural ubiquitin to a similar resolution using a Siemens-Stoe AED-24-circle diffractometer and graphite-monochromated, Cu-Kα radiation from a sealed beam tube operating at 50 kV, 35 mA. 4295 reflections were collected.

Data Set Comparison

Comparison of the synthetic and natural ubiquitin structure factors with those from the published model for natural ubiquitin and 58 associated water molecules (1UBQ in the Brookhaven Databank), revealed that the data collected for the synthetic protein conform to the model somewhat better than the data collected from the natural ubiquitin crystal; Rsyn-mod = 18.3% and Rnat-mod = 19.8%. After model temperature factor refinement, the R-factor dropped to 17.1% for the synthetic data set, which is close to the final R-index for the original model of 17.6%.

To reduce the influence of the model upon the phases, and so reveal possible real minor differences between native and synthetic ubiquitin, experimental data sets were compared directly rather than each of them to the model structure factors. Rsyn-nat was found to be 10.4% confirming the high overall similarity of the

structures. This value is somewhat higher than expected but reflects the comparison of a data-set collected serially (natural) with one collected on an area detector.

To estimate the size of differences, (2[Fsyn]-[Fnat]) Fourier synthesis and difference ([Fsyn]-[Fnat]) synthesis were compared. The phases derived from the natural ubiquitin model before refinement with the data collected in this work, were used. The maximum of the difference map was found to be 5 times lower than r.m.s. deviation of the first synthesis suggesting that the synthetic and natural protein structures are indistinguishable at this resolution.

Model Refinement

The molecular model of native ubiquitin was taken from the Brookhaven databank and used as a starting model. Careful inspection of the (2[Fsyn-Fnat]) density map revealed that the conformation of some of the side-chains, and the position of some of the water molecules could be improved. The model was therefore refined using release 4 of the TNT restrained least-squares procedure. 4349 reflections were used for TNT refinement and the set of 5031 reflections including poorly measured weak reflections was used for electron density calculation in the resolution range 6.0 to 1.8 A^{o}. After every round of refinement (10 cycles), the (2[Fo]-[Fc]) and difference Fourier maps were inspected and adjusted using the program FRODO running on an ESV 10/20 molecular graphics workstation. Water molecules were added during this process. Because the last 3 residues at the C-terminus are very flexible and therefore poorly defined, throughout the refinement their positions and occupancies were fixed as in the original model. The positions of the 2 C-terminal residues previously determined were based on the difference map between the whole ubiquitin molecule and that with the last 2 removed since the calculated phases are more reliable than if based on the data collected from ubiquitin-76 only.

Each round of refinement consisited of 5 cycles of positional refinement for the non-hydrogen atoms in the protein and water molecules, followed by 5 subsequent cycles of temperature factor and occupancy refinement. All positive peaks in the (2[Fo]-[Fc]) and ([Fo]-[Fc]) Fourier maps higher than the 1σ level that were not already covering model atoms were examined. Water molecules were added to the model only if they occupied stereochemically reasonable positions. After 6 rounds of refinement the model had converged to R=16.5%.

1. J.A. Ellman, D. Mendel and P.G. Schultz, *Science*, 1992, **255**, 197.
2. J. Schneider and S.B.H. Kent, *Cell*, 1988, **54**, 363.
3. S. Vijay-Kumar, C.E. Bugg and K.D. Wilkinson, W.J. Cook, *Proc. Natl. Acad.Sci. USA*, 1985, **82**, 3582.
4. P.L. Weber and S.C. Brown, L. Mueller, *Biochemistry*, 1987, **26**, 7282.
5. D.L. Distefano and A.J. Wand, *Biochemistry*, 1987, **26**, 7272.
6. A. Hersko and A. Ciechanover, *Ann. Rev. Biochem.*, 1982, **251**, 335.
7. G. Goldstein, M. Scheid, U. Hammerling, E.A. Boyse, D.H. Schlesinger and H.D. Niall, *Proc. Natl. Acad. Sci. USA*, 1975, **72**, 11.
8. O. Wiborg, M.S. Pedersen, A. Wind, L.E. Bergland and K.A. Marcker, *EMBO. J.*, 1985, **4**, 755.
9. U. Bond and M.J. Schlesinger, *Mol. Cell. Biol.*, 1985, **5**, 949.
10. E. Ozkaynak, D. Finley and A. Varshavsky, *Nature*, 1984, **312**, 663.
11. E. Dworkin-Rastl, A. Shrotkowski and M.B. Dworkin, *Cell*, 1984, **39**, 321.
12. K. Gausing and R. Barkandottir, *Eur. J. Biochem*, 1986, **158**, 57.
13. A.N. Mayer and K.D. Wilkinson, *Biochemistry*, 1989, **28**, 166.
14. P.K. Lund, B.M. Moats-Staats, J.G. Simmons, E. Hoyt, A.J. D'ercole, F. Martin and J.J. Van Wyk, *J. Biol. Chem*, 1985, **260**, 394.
15. G. Salvesen, C. Lloyd and D. Farley, *Nucleic Acids Res.*, 1987, **15**, 5485.
16. A. Muller-Taubenberger, M. Westphal, E. Jaeger, A. Noegel and G. Gerisch, *F.E.B.S.* , 1988, **229**, 273.
17. E. Ozkaynak, D. Finley, M.J. Solomon and A. Varashavsky, *EMBO. J.*, 1987, **6**, 1429.
18. B.P. Monia, D.J. Ecker, S. Jonnalagadda, J. March, L. Gotlib, T.R. Butt and S.T. Crooke, *J. Biol. Chem.*, 1989, **264**, 4093.
19. I.L. Golgknopf and H. Busch, *Proc. Natl. Acad. Sci. USA.*, 1977, **74**, 864.
20. L.T. Hunt and M.O. Hayhoff, *Biochem. Biophys. Res. Commun.*, 1977, **74**, 650.
21. M.H.P. West and W.M. Bonner, *Nucleic Acids Res.*, 1980, **8**, 4671.
22. B. Pina and P. Suau, *Biochem. Biophys. Res. Commun.*, 1985, **133**, 505.
23. M.H.P. West and W.M. Bonner, *Biochemistry*, 1980, **19**, 3238.
24. S.C. Albright, P.P. Nelson and W.T. Garrard, *J. Biol. Chem.*, 1979, **254**, 1065.
25. B.E. Nickel and J.R. Davie, *Biochemistry*, 1987, **28**, 964.
26. J.R. Davie, G.P. Delcuve, B.E. Nickel, R. Moyer and G. Bailey, *Cancer Res.*, 1987, **47**, 5407.

27. I.L. Golgknopf, G. Watson, N.R. Ballal and H. Busch, *J. Biol. Chem.*, 1980, **255**, 10555.
28. L. Levinger and A. Varashavsky, *Cell*, 1982, **28**, 375.
29. A. Varashavsky, L. Levenger, O. Sundin, J. Barsoum, E. Ozkaynak, P.S. Swerdlow and D. Findley, *Cold Spring Harbour Symp. Quant. Biol.*, 1982, **47**, 511.
30. J. Barsoum and A. Varashavsky, *J. Biol. Chem.*, 1985, **260**, 7688.
31. C. Ericsson, I.L. Goldknopf and B. Daneholt, *Exp. Cell Res.*, 1986, **167**, 127.
32. S. Matsui, B.K. Seon and A.A. Sandberg, *Proc. Natl. Acad. Sci. USA.*, 1979, **76**, 6386.
33. M. Glotzer, A.W. Murray and M.W. Kirschner, *Nature*, 1991, **349**, 132.
34. M. Siegelman, M.W. Bond, W.M. Gallatin, T. St John, H.T. Smith, V.A. Fried and I.L. Weissman, *Science*, 1986, **231**, 829.
35. W.M. Gallatin, T. St John, M. Siegelman, R. Reichert, E.C. Butcher and I.L. Weissman, *Cell*, 1986, **44**, 673.
36. T. St. John, W.M. Gallatin, M. Siegelman, H.T. Smith, V.A. Fried and I.L. Weissman, *Science*, 1986, **231**, 719.
37. Y. Yarden, J.A. Escobedo, W-J. Kuang, T.L. Yang Feng, T.O. Daniel, P.M. Tremble, E.Y. Chen, M.E. Ando, R.N. Harkins, U. Francke, V.A. Fried, A. Ullrich and L.T. Williams, *Nature*, 1986, **323**, 226.
38. D.W. Leung, S.A. Spenser, G. Cachianes, R.G. Hammonds, C. Collins, W.J. Henzel, R. Barnard, M.J. Waters and W.I. Wood, *Nature*, 1987, **330**, 537.
39. T. Okabe, M. Fujisawa, A. Mihara, S. Sato, N. Fujiyoshi and F. Takaku, *J. Cell Biol.*, 1986, **103**, 442.
40. (a) M. Kidd, *Nature*, 1963, **197**, 192.
    (b) H. Wisniewski, H.K. Narang and R.D. Terry, *J. Neurol. Sci.*, 1976, **27**, 173.
41. H. Mori, J. Kondi and Y. Ihara, *Science*, 1987, **235**, 1641.
42. J. Lowe and R.J. Mayer, *Neuropath. Appl. Neurobiol.*, 1990, **16**, 281.
43. I.M. Schwartz, L. Kosz and B.K. Kay, *Proc. Natl. Acad. Sci. USA.*, 1990, **87**, 6594.
44. D. Finley and V. Chau, *Ann. Rev. Cell Biol.*, 1991, **7**, 25.
45. M. Rechsteiner, *Ubiquitin*, 1988, Plenum Publishing Corp., New York.
46. A. Ciechanover, Y. Hod and A. Hershko, *Biochem. Biophys. Res. Commun.*, 1978, **81**, 1100.
47. K.D. Wilkinson, M.K. Urban and A.L.Haas, *J. Biol. Chem.*, 1980, **255**, 7529.

48. A. Hershko and A. Ciechanover, *Ann. Rev. Biochem.*, 1982, **251**, 335.
49. A. Hershko, A. Ciechanover, H. Heller, A.L. Haas and I.A. Rose, *Proc. Natl. Sci. USA.*, 1980, **77**, 1783.
50. A. Ciechanover, H. Hellor, R. Katz-Etzion and A. Hershko, *Proc. Natl. Sci. USA.*, 1981, **78**, 761.
51. A.L. Haas, J.V.B. Warms, A. Hershko and I.A. Rose, *J. Biol. Chem.*, 1982, **257**, 2543.
52. A. Hershko, H. Heller, S. Elias and A. Ciechanover, *J. Biol. Chem.*, 1983, **258**, 8206.
53. A. Ciechanover, S. Elias, S. Heller and A. Hershko, *J. Biol. Chem.*, 1982, **257**, 2537.
54. E. Zachsenhaus and R. Sheinin, *EMBO. J.*, 1990, **9**, 2923.
55. P.M. Handley, M. Muechler, N.R. Siegel, A. Ciechanover and A.L. Schwartz, *Proc. Natl. Acad. Sci. USA.*, 1991, **88**, 258.
56. T. Eki, T. Enomoto, A. Miyajima, H. Miyazawa and Y. Murakami, *J. Biol. Chem.*, 1990, **265**, 26.
57. J.P. McGrath, S. Jentsch and A. Varashavsky, *EMBO. J.*, 1991, **10**, 227.
58. P.M. Hatfield and R.D. Viestra, *Biochemistry*, 1989, **28**, 735.
59. S. Jentsch, W. Seufert, T. Sommer and H.A. Rains, *Trends. Biochem. Sci.*, 1990, **15**, 195.
60. M.L. Sullivan and R.D. Viestra, *Proc. Natl. Acad. Sci. USA.*, 1989, **86**, 9861.
61. S. Jentsch, J.P. McGrath and A. Varashavsky, *Nature*, 1987, **329**, 131.
62. A.L. Haas, P.M. Rebach, G. Pratt and M. Rechsteiner, *J. Biol. Chem.*, 1990, **265**, 21664.
63. A. Morrison, E.J. Miller and L. Prakash, *Mol. Cell Biol.*, 1988, **8**, 1179.
64. Y. Reiss and A. Hershko, *J. Biol. Chem.*, 1990, **265**, 3685.
65. H. Heller and A. Hershko, *J. Biol. Chem.*, 1990, **265**, 6532.
66. B. Bartel, I. Wunning and A. Varashavsky, *EMBO. J.*, 1990, **9**, 3179.
67. Y. Reiss, D. Kaim and A. Hershko, *J. Biol. Chem.*, 1988, **263**, 2693.
68. S. Jonnalagadda, T.R. Butt, B.P. Monia, C.K. Mirabelli and L. Gotlid, *J. Biol. Chem.*, 1989, **264**, 10637.
69. A. Hershko, E. Leshinsky, D. Ganoth and H. Heller, *Proc. Natl. Acad. Sci. USA.*, 1984, **81**, 1619.
70. R. Hough, G. Pratt and M. Rechsteiner, *J. Biol. Chem.*,1986, **261**, 2400.
71. L. Waxman, J.M. Fagan and A.L. Goldberg, *J. Biol. Chem.*, 1987, **262**, 8303.

72. D. Ganoth, E. Lehinsky, E. Eytan and A. Hersko, *J. Biol. Chem.*, 1988, **263**, 12412.
73. E. Eytan, D. Ganoth, T. Armon and A. Hershko, *Proc. Natl. Acad. Sci. USA.*, 1989, **86**, 7751.
74. K. Tanaka, and A. Ichihara, *F.E.B.S. Lett.*, 1988, **236**, 159.
75. J. Driscoll, J. Frydman and A.L. Goldberg, *J. Biol. Chem.*, Paper Submitted.
76. T. Hadari, J.V.B. Warms, I.A. Rose and A. Hershko, *J. Biol. Chem.*, 1992, **267**, 719.
77. M. Rechsteiner, *Ann. Rev. Cell. Biol.*, 1987, **3**, 1.
78. A. Hershko, E. Eytan, A. Ciechanover and A.L. Haas, *J. Biol. Chem.*, 1982, **257**, 13964.
79. F.J. Docherty, N.U. Osborn, J.A. Wassel, P.E. Heggie and L. Laszlo, *J. Biochem.*, 1989, **263**, 47.
80. S. Bigelow, R. Hough and M. Rechsteiner, *Cell*, 1981, **25**, 83.
81. Gropper. R, R.A. Brandt, S. Elias, C.F. Bearer, A. Mayer, A.L. Schwartz and A. Ciechanover, *J. Biol. Chem.*, 1991, **266**, 3602.
82. J.L. Brown and W.K. Roberts, *J. Biol. Chem.*, 1976, **251**, 1009.
83. A. Mayer, N.R. Siegel, A.L. Schwartz and A. Ciechanover, *Science*, 1989, **244**, 1480.
84. H. Goner, A.L. Schwartz and A. Ciechanover, *J. Biol. Chem.*, 1991, **266**, 19221.
85. A. Bachmair, D. Finley and A. Varashavsky, *Science*, 1986, **234**, 179.
86. D.K. Gonda, A. Bachmair, I. Wunning, J.W. Tobias, W.S. Lone and A. Varashavsky, *J. Biol. Chem.*, 1989, **264**, 16700.
87. Y. Reiss and A. Hershko, *J. Biol. Chem.*, 1990, **265**, 3685.
88. S. Ferber and A. Ciechanover, *Nature*, 1987, **326**, 808.
89. A. Ciechanover, S. Ferber, D. Ganoth, S. Elias and A. Hershko, *J. Biol. Chem.*, 1988, **263**, 11155.
90. A. Hershko and H. Heller, *Biochem. Biophys. Res. Commun.*, 1985, **128**, 1079.
91. U. Chau, J.W. Tobias, A. Bachmair, D. Marriot and D.J. Ecker, *Science*, 1989, **243**, 1576.
92. A. Ciechanover, D. Finley and A. Varashavsky, *Cell*, 1984, **37**, 57.
93. A. Ciechanover, D. Finley and A. Varashavsky, *Cell*, 1984, **37**, 43.
94. Q. Devaraux, R. Wells and M. Rechsteiner, *J. Biol. Chem.*, 1990, **265**, 6323.
95. H.A. Parag, B. Raboy and R.G. Kulka, *EMBO. J.*, 1987, **6**, 55.

96. D. Finley and A. Varashavsky, *Trends Biochem. Sci.*, 1985, **10**, 343.
97. I.A. Rose and J.V.B. Warms, *Proc. Natl. Acad. Sci. USA.*,1987, **84**, 1477.
98. D. Finley, E. Ozkaynak and A. Varashavsky, *Cell*, 1987, **48**, 1035.
99. W. Seufert and S. Jentsch, *EMBO. J.*, 1990, **9**, 543.
100. P.M. Kelly and M.J. Schlesinger, *Mol. Cell Biol.*, 1982, **2**, 267.
101. K.L. Redman and M. Rechsteiner, *J. Biol. Chem.*, 1988, **263**, 4926.
102. S.C. Harrison, *Nature*, 1991, **353**, 715.
103. K.L. Redman and M. Rechsteiner, *Nature*, 1989, **338**, 438.
104. D. Finley, B. Bartel and A. Varashavsky, *Nature*, 1989, **338**, 394.
105. B.P. Monia, D.J. Ecker and S.T. Crooke, *Biotechnology*, 1990, **8**, 209.
106. D. Ecker, J.M. Stadel, T.R. Butt, J.A. March, B.P. Monia, D.A. Powers, J.A. Gorman, P.E. Clark, F. Warren, A. Shatzman and S.T. Crooke, *J. Biol. Chem.*,1989, **264**, 7715.
107. T.R. Butt, S. Jonnalagadda, B.P. Monia, E.J. Sternberg, J. March, J.M. Stadel, D.J. Ecker and S.T. Crooke, *Proc. Natl. Acad. Sci. USA.*, 1989, **86**, 2540.
108. B.W. Cherney, B. Chaudry, T.R. Butt and M. Smulson, *Biochemistry*, 1991, **30**, 10420.
109. K.D. Wllkinson, M.J. Cos, L.B. O'Conner and R. Shapiro, *Biochemistry*, 1986, **25**, 4999.
110. R.E. Lenkinski, D.M. Chen, J.D. Glickson and G. Goldstein, *Biochim. Biophys. Acta.*, 1977, **494**, 126.
111. D. Schlesinger and G. Goldstein, *Nature*, 1975, **255**, 423.
112. J. Jenson, G. Goldstein and E. Breslow, *Biochim. Biophys. Acta.*, 1980, **624**, 378.
113. K. Ugurbil and R. Bersohn, *Biochemistry*,1977, **16**, 895.
114. W.J. Cook, F.L. Suddath, C.E. Bugg and G. Goldstein, *J. Mol. Biol.*,1979, **130**, 353.
115. S. Vijay-Kumar, C.E. Bugg and W.J. Cook, *J. Mol. Biol.*, 1987, **194**, 531.
116. S. Vijay-Kumar, C.E. Bugg, K.D. Wilkinson, R.D. Vierstra and P.M. Hatfield, *J. Biol. Chem.*, 1987, **262**, 6396.
117. J. Jeener, B.H. Meier, P. Bachmann and R.R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546.
118. K. Wuthrich, G. Wider, G. Wagner and W. Braun, *J. Mol. Biol.*,1982, **155**, 311.
119. M. Billeter, W. Braun and K. Wuthrich, *J. Mol. Biol.*,1982, **155**, 321.
120. K. Wuthrich, *Biopolymers*, 1983, **22**, 131.

121. K. Wuthrich, *NMR of Proteins and Nucleic Acids*, 1986, Wiley, New York.
122. A. Shaka and R. Freeman, *J. Magn. Reson.*, 1983, **51**, 169.
123. P.L. Weber and L.M. Mueller, *J. Magn. Reson.*, 1987, **73**, 184.
124. L. Braunschweiler and R.R. Ernst, *Mol. Phys.*, 1983, **53**, 521.
125. S. Macura and R.R. Ernst, *Mol. Phys.*, 1980, **41**, 95.
126. M. Karplus, *J. Chem. Phys.*, 1959, **30**, 11.
127. P.L. Weber, D.J. Ecker, J. March, S.T. Crooke and L. Mueller, *Trans. Amer. Cryst. Assoc.*, 1988, **24**, 91.
128. S.J. Weiner, P.A. Kollman, D.A. Case, V.C. Singh, C. Ghio, G. Alagona, S. Profeta Jr. and P. Weiner, *J. Am. Chem. Soc.*, 1984, **106**, 765.
129. D.R. Hare and R.D. Morrison, *Hare Research Inc.*, 14810, 216th NE, Woodinville, WA 98071.
130. D.J. Ecker, T.R. Butt, J. Marsh, E.J. Sternberg, N. Margolis, B.P. Monia, S. Jonnalagadda, M.I. Khan, P.L. Weber, L. Mueller and S.T. Crooke, *J. Biol. Chem.*, 1987, **262**, 14213.
131. K.D. Wilkinson and A.N. Mayer, *Arch. Biochem. Biophys.*, 1986, **250**, 390.
132. S.N. Timasheff and H. Inove, *Biochemistry*, 1968, **7**, 2501.
133. T. Arakawa and D. Doddette, *Arch. Biochem. Biophys.*, 1985, **240**, 21.
134. M.M. Harding, D.H. Williams and D.N. Woolfson, *Biochemistry*, 1991, **30**, 3120.
135. D.J. Ecker, T.R. Butt, J. March, E. Sternberg, A. Shatzman, J.S. Dixon, P.L. Weber and S.T. Crooke, *J. Biol. Chem.*, 1989, **264**, 1887.
136. S. Bamezai, M.A.T. Banez and E. Breslow, *Biochemistry*, 1990, **29**, 5389.
137. A.E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci. USA.*, 1936, **22**, 439.
138. W. Kauzmann, *Adv. Protein. Chem.*, 1959, **14**, 1.
139. C. Tanford, *J. Am. Chem. Soc.*, 1962, **84**, 4240.
140. A.E. Eriksson, W.A. Baase, X-J. Zhang, D.W. Heinz, M. Blaber, E.P. Baldwin and B.W. Matthews, *Science*, 1992, **255**, 178.
141. J.T. Kellis Jr., K. Nyberg and A.R. Fersht, *Nature*, 1988, **333**, 784.
142. J.T. Kellis Jr., K. Nyberg and A.R. Fersht, *Biochemistry*, 1989, **28**, 4914.
143. D.H. Williams, J.P.L. Cox, A.J. Doig, M. Gardner, U. Gerhard, P.T. Kaye, A.R. Lal, I.A. Nicholls, C.J. Salter and R.C. Mitchell, *J. Am. Chem. Soc.*, 1991, **113**, 7020.
144. A.J. Doig and D.H. Williams, *J. Am. Chem. Soc.*, 1992, **114**, 338.
145. B.A. Shirley, P. Stanssens, U. Hahn and C.N. Pace, *Biochemistry*, 1992, **31**, 725.

146. W.A. Lim and R.T. Sauer, *J. Mol. Biol.*, 1991, **219**, 359.
147. F.K. Lau and K.A. Dill, *Proc. Natl. Acad. Sci. USA.*, 1990, **87**, 638.
148. P.S. Kim and R.L. Baldwin, *Ann. Rev. Biochem.*, 1990, **59**, 631.
149. H. Roder and K. Wuthrich, *Proteins*, 1986, **1**, 34.
150. H. Roder, G.A. Elove and S.W. Englander, *Nature*, 1988, **335**, 700.
151. M. Bycroft, A. Matouschek, J.T. Kellis Jr., L. Serrano and A.R. Fersht, *Nature*, 1990, **346**, 488.
152. A. Miranker, S.E. Radford, M. Karplus and C.M. Dobson, *Nature*, 1991, **349**, 633.
153. J.B. Vdgaonkar and R.L. Baldwin, *Nature*, 1988, **335**, 649.
154. K. Kuwajima, *Proteins*,1989, **6**, 87.
155. M.S. Briggs and H. Roder, *Proc. Natl. Acad. Sci. USA.*, 1992, **89**, 2017.
156. E. Fisher and E. Fourneau, *Ber. Dtsch. Chem. Ges.*1901, **34**, 2868.
157. R.B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149.
158. M. Bodanszky and J. Martinez, *In The Peptides*, Vol 5, (E. Gross, J. Meinhofer, Eds.), p111, New York Academic Press, 1983.
159. M. Bodanszky, *Peptide Chemistry*, Springer-Verlag, Berlin, 1988 ed.
160. M. von Brenner and W. Hofer, *Helv. Chim. Acta*, 1961, **215**, 1794.
161. E. Kisher, *Ber. Dtsch. Chem. Ges.*, 1903, **36**, 2094.
162. T. Curtius, *Ber. Dtsch. Chem. Ges.*, 1902, **35**, 3226.
163. D. Sarantakis, J. Teichman, E. Lien and R. Fenichel, *Biochem. Biophys. Res. Commun.*, 1976, **73**, 336.
164. Y. Kiso, M. Satomi, K. Ukawa and T. Akista, *J. Am. Chem. Soc. Commun.*, 1980, 1063.
165. (a) J.R. Vaughan Jr., *J. Am. Chem. Soc.*, 1951, **73**, 3547.
(b) R. Ramage, D. Hopton, M.J. Parrot, R.S. Richardson, G.W. Kenner and G.A. Moore, *J. Chem. Soc. Perkin Trans. 1*, 1985, 461.
166. M. Bodanszky, *Nature*, 1955, **175**, 685.
167. W. Konig and R. Geiger, *Chem. Ber.*, 1970, **103**, 2034.
168. L. Kisfaludu and I. Schon, *Synthesis*, 1986, 303.
169. M. Bergman and L. Zervas, *Ber. Dtsch. Chem. Ges.*, 1932, **65**, 1192.
170. F.C. McKay and N.F. Albertson, *J. Am. Chem. Soc.*, 1957, **79**, 4686.
171. P. Sieber and B. Iselin, *Helv. Chem. Acta*, 1968, **51**, 614.
172. L.A. Carpino and G.Y. Han, *J. Am. Chem. Soc.*, 1970, **92**, 5748.
173. S.B.H. Kent, *Annu. Rev. Biochem.*, 1988, **57**, 957.
174. K. Noda, S. Tereda and N. Izumiya, *Bull. Chem. Soc. Jpn.*, 1969, **43**, 1883.

175. R. Ramage and J. Green, *Tetrahedron Lett.*, 1987, **28**, 2287.
176. S. Mosjov, *J. Org. Chem.*, 1980, **41**, 1032.
177. R. Columbo, F. Columbo and J.H. Jones, *J. Chem. Soc. Chem. Commun.*, 1984, 292.
178. P. Sieber and B. Riniker, *Tetrahedron Lett.*, 1987, **28**, 6031.
179. T. Brown and J.H. Jones, *J. Chem. Soc. Chem. Commun.*, 1981, **648**.
180. S. N. McCurdy, *Pept. Res.*, 1989, **2**, 147.
181. C.R. Wu, J.D. Wade and G.W. Tregear, *Int. J. Pept. Prot. Res.*, 1988, **31**, 47.
182. P. Sieber and B. Riniker, *Tetrahedron Lett.*, 1991, **32**, 739.
183. V. du Vigneaud, C. Ressler, J.M. Swan, C.W. Roberts, P.G. Katsoyannis and S. Gordon, *J. Am. Chem. Soc.*, 1953, **75**, 4879.
184. M. Bodanszky, M.A. Ondetti, S.D. Levine and N.J. Williams, *J. Am. Chem. Soc.*, 1967, **89**, 6753.
185. R.B. Merrifield, *J. Am. Chem. Soc.*, 1964, **86**, 304.
186. S.S. Wang, *J. Am. Chem. Soc.*, 1973, **95**, 1328.
187. R. Ramage, J. Green and O.M. Ojunjobi, *Tetrahedron Lett.*, 1987, **30**, 2149.
188. J-P. Briand, A. Van Dorsselaer, B. Raboy and S. Muller, *Pept. Res.*, 1989, **2**, 381.
189. C.D. Chang and J. Meienhofer, *Int. J. Pept. Prot. Res.*, 1978, **11**, 246.
190. J.H. Jones, W.I. Ramage and M.J. Witty, *Int. J. Pept. Prot. Res.*,1980, **15**, 301.
191. K. Shaw, University of Edinburgh, Personal Communcation.
192. V.K. Sarin, S.B. Kent, J.P. Tam and R.B. Merrifield, *Anal. Biochem.*, 1981, 117.
193. Performed by K. Kimber of Vestec Interon on a VT2000 Laser Desorption T.O.F Mass Spectrometer, Unit 5, Britannia Rd., Sale, Cheshire.
194. M. Bodanszky, J. Martinez, *Synthesis*, 1981, 6068.
195. Performed by the VG BIOTECH (Fisons) on a BIO-Q Electrospray Mass Spectrometer.
196. M.J. Cox, R. Shapiro and K.D. Wilkinson, *Anal. Biochemistry.*1986, **154**, 345.
197. C. Tanford and G.L. Roberts Jr., *J. Am. Chem. Soc.*, 1952, **74**, 2509.
198. C. Tanford and J. Epstein, *J. Am. Chem. Soc.*, 1954, **76**, 2163.
199. C. Tanford, J.D. Hauenstein and D.G. Rands, *J. Am. Chem. Soc.*, 1956, **77**, 6409.
200. S.S. Lehrer and P.C. Leavis, *Methods Enzymol.*, 1978, **49**, 222.

201. C.D. Lindsay and R.H. Pain, *J. Biochem.*, 1990, **192**, 133.
202. K. Kash, *J. Chem. Phys.*1952, **20**, 71.
203. C.N. Pace, *Methods Enzymol.*, 1896, **131**, 266.
204. Performed by N. Price and S. Kelly, Biochemistry Dept., University of Stirling.
205. S.W. Provencher and J. Glockner, *Biochemistry*, 1981, **20**, 33.
206. C.T. Chang, C.S.C. Wu and J.T. Yang, *Anal. Biochem.*, 1978, **91**, 13.
207. A.E. Derome, *Modern NMR Techniques for Chemical Research*, 1987, Vol. 6, Permagon Press.
208. Performed by D. Alexeev, S.M. Fawcett, M.A. Turner and L. Sawyer, Biochemistry Dept., University of Edinburgh.
209. T.A. Jones, *J. Appl. Cryst.*, 1978, **11**, 268.
210. Performed by J. Mayer and J. Arnold, Biochemistry Dept., Queens Medical Centre, University of Nottingham.
211. W.A. Lim and R.T. Sauer, *Nature*, 1989, **339**, 31.
212. M.J. Bogusky, P. Leighton, R.A. Schiksnin, A. Khoury, P. Lu and S.J. Opella, *J. Mag. Res.*, 1990, **80**, 11.
213. A. Chakrabartty, J.A. Schellman and R.L. Baldwin, *Nature*, 1991, **351**, 586.
214. L. Serrano, J-L, Neira, J. Sancho and A.R. Fersht, *Nature*, 1992, **356**, 453.
215. R.E. Shute and D.H. Rich, *Synthesis*, 1987, 346.
216. D.W. Thomas, University of Edinburgh, Personal Communcation.
217. C.M. Dupureur, Y. Li and M-D. Tsai, *J. Am. Chem. Soc.*, 1992, **114**, 2748.
218. P.Y. Chou and G.D. Fasman, *Biochemistry*, 1974, **13**, 211.
219. J. Garnier, D.J. Osguthorpe and B. Robson, *J. Mol. Biol.*, 1978, **120**, 97.
220. N.P. Pavletich and C.O. Pabo, *Science*, 1991, **252**, 809.
221. J.G. Omnichinski, G.M. Clore, E. Appella, K. Sakagushi and A.M. Gronenborn, *Biochemistry*, 1990, **29**, 9324.
222. G.L. Ellman, *Arch. Biochem. Biophys.*, 1959, **82**, 70.
223. Performed by Welmet Protein Characterization Group, Biochemistry Dept., University of Edinburgh.
224. Performed by A. Keane of Finnigan Mat on a LASERMAT Laser Desorption T.O.S. Mass Spectrometer.
225. Performed by R. Layfield, Biochemistry Dept., Queens Medical Centre, University of Nottingham.
226. A.D. Frankel, J.M. Berg and C.O. Pabo, *Proc. Natl. Acad. Sci. USA.*, 1987, **84**, 4841.

227. G. Parraga, S. Harvath, L. Hood, E.T. Young and R.E. Klevit, *Proc. Natl. Acad. Sci. USA.*, 1990, **87**, 137.

228. Y. Nozaki, *Methods Enzymol.*, 1972, **26**, 43.

Lectures Attended

Organic Research Seminars (various Speakers).

Recent Advances in Organic Chemistry (various speakers, University of Edinburgh).

Medicinal Chemistry (Professor R. Baker and Dr. P. Lesson, Merck Sharp and Dohme).

Aspects and Applications of NMR Spectroscopy (various speakers, University of Edinburgh).

Protein Engineering (various speakers, Royal Society of Edinburgh).

Departmental German Course (passed 1990).

Royal Society of Chemistry, Bioorganic Group One-Day Postgraduate Symposium (various speakers, Warwick University).

Royal Society of Chemistry, Pre-Doctoral Symposium (various speakers, Trinity College Dublin).

Annual Bioorganic Symposium (various speakers, Gregynog).

Scottish Protein Group (various speakers, Heriot-Watt University).